PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q-1/68

A2

(11) International Publication Number: WO 98/53097

(43) International Publication Date: 26 November 1998 (26.11.98)

(21) International Application Number: PCT/CA98/00488

(22) International Filing Date: 21 May 1998 (21.05.98)

(30) Priority Data:
08/861,774 22 May 1997 (22.05.97) US

(71) Applicant: TERRAGEN DIVERSITY INC. [CA/CA]; University of British Columbia, Suite 300, 2386 East Mall, Vancouver, British Columbia V6T 1Z3 (CA).

(72) Inventors: WATERS, Barbara; 5706 Timbervalley Road, Delta, British Columbia V4L 2E6 (CA). MIAO, Vivian, P., W.; 13750 31 Avenue, Surrey, British Columbia V4P 2B7 (CA). YAP, Wai, Ho; 5 Elite Terrace, Singapore 458748 (SG). SEOW, Kah, Tong; 8 Jln Aneka, Serene Park, Johor Baru, Johor 80300 (MY).

(74) Agent: DEETH WILLIAMS WALL; National Bank Builing, Suite 400, 150 York Street, Toronto, Ontario M5H 3S5 (CA).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR BIOACTIVE MOLECULES

(57) Abstract

Degenerate primers which hybridize with various classes of antibiotic biosynthesis gene were used to amplify fragments of DNA from soil and lichen extracts. Cloning and sequencing of the amplified products showed that these products included a variety of novel and previously uncharacterized antibiotic biosynthesis gene sequences, the products of which have the potential to be active as antibiotics, immunosuppressors, antitumor agents, etc. Thus, antibiotic biosynthesis genes can be recovered from soil or lichens by combining a sample with a pair of amplification primers under conditions suitable for polymerase chain reaction amplification, wherein the primer set is a degenerate primer set selected to hybridize with conserved regions of known antibiotic biosynthetic pathway genes, for example Type I and Type II polyketide synthase genes, isopenicillin N synthase genes, and peptide synthetase genes, cycling the combined sample through a plurality of amplification cycles to amplify DNA complementary to the primer set; and isolating the amplified DNA.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR BIOACTIVE MOLECULES

DESCRIPTION

BACKGROUND OF THE INVENTION

This application relates to a method for the isolation of biosynthesis genes for antibiotics and other bioactive molecules from complex natural sources such as humus, soil and lichens.

Antibiotics play an important role in man's efforts to combat disease and other economically detrimental effects of microorganisms. Traditionally, antibiotics have been identified by screening microorganisms, especially those found naturally in soil, for their ability to produce an antimicrobial substance. In some cases, the gene or genes responsible for antibiotic synthesis have then been identified and cloned into producer organisms which produce the antibiotic in an unregulated manner for commercial applications. However, it has been estimated that less than 1% of the microorganisms present in soil are culturable. Torsvik et al., *Appl. Environ. Microbiol.* 56: 782-787 (1990). Thus, much of the genetic diversity potentially available in soil microorganisms is unavailable through traditional techniques.

As pathogenic microorganisms become increasingly resistant to known antibiotics, it would, however, be highly desirable to be able to access the reservoir of genetic diversity found in soil, and to facilitate the exploration of new species of antibiotics which may be made by the vast numbers of unculturable organisms found there. It would further be desirable to have access to novel biosynthetic enzymes and the genes encoding such enzymes, which could be used in recombinant organisms for antibiotic production or for *in vitro* enzymatic synthesis of desirable compounds. Thus, it is an object of the present invention to provide a method and compositions for isolating DNA and DNA fragments encoding enzymes relevant to the production of pharmaceutically active molecules such as antibiotic biosynthesis enzymes.

SUMMARY OF THE INVENTION

We have now identified degenerate primers which hybridize with various classes of antibiotic biosynthesis genes, and have used such primers to amplify fragments of DNA from soil and lichen extracts. Cloning and sequencing of the amplified products showed that these products included a variety of novel and previously uncharacterized antibiotic biosynthesis gene sequences, the products of which have the potential to be active as antibiotics, immunosuppressors, antitumor agents, etc. Thus, antibiotic biosynthesis genes can be recovered from soil by a method in accordance with the present invention comprising the steps of:

(a) combining a soil-derived sample with a pair of amplification primers under conditions suitable for polymerase chain reaction amplification, wherein the primer set is a degenerate primer set selected to hybridize with conserved regions of known antibiotic biosynthetic pathway genes, for example Type I and Type II polyketide synthase genes, isopenicillin N synthase genes, and peptide synthetase genes;

(b) cycling the combined sample through a plurality of amplification cycles to amplify DNA complementary to the primer set; and

(c) isolating the amplified DNA.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, antibiotic biosynthesis genes can be recovered from soil and lichens by a method comprising the steps of:

(a) combining a humic or lichen-derived sample with a pair of amplification primers under conditions suitable for polymerase chain reaction amplification, wherein the primer set is a degenerate primer set selected to hybridize with conserved regions of an antibiotic biosynthesis gene;

(b) cycling the combined sample through a plurality of amplification cycles to amplify DNA complementary to the primer set; and

(c) isolating the amplified DNA.

As used in the specification and claims of this application, the term "humic or lichen-derived sample" encompasses any sample containing the DNA found in lichens or in samples of humic materials including soil, mud, peat moss, marine sediments, and effluvia

from hot springs and thermal vents in accessible form for amplification, substantially without alteration of the natural ratios of such DNA in the sample. One exemplary form of a humic sample is a sample obtained by performing direct lysis as described by Barns et al., *Proc. Nat'l Acad. Sci. USA* 91:1609-1613 (1994) on a soil sample and then purifying the total DNA extract by column chromatography. Related extraction methods can be applied to the isolation of community DNA from other environmental sources. See, Trevors et al., eds. *Nucleic Acids in the Environment*, Springer Lab Manual (1995). Lichen-derived samples may be prepared from foliose lichens by the method of fungal DNA extraction described by Miao et al., *Mol. Gen. Genet.* 226: 214-223 (1991). Specific non-limiting procedures for isolation of DNA from humic and lichen samples are set forth in the examples herein.

The humic or lichen-derived sample is combined with at least one, and optionally with several pairs of amplification primers under conditions suitable for polymerase chain reaction amplification. Polymerase chain-reaction (PCR) amplification is a well known process. The basic procedure, which is described in US Patent No. 4,683,202 and 4,683,195, which are incorporated herein by reference, makes uses of two amplification primers each of which hybridizes to a different one of the two strands of a DNA duplex. Multiple cycles of primer extension using a polymerase enzyme and denaturation are used to produce additional copies of the DNA in the region between the two primers. In the present invention, PCR amplification can be performed using any suitable polymerase enzyme, including Taq polymerase and Thermo Sequenase™.

The amplification primers employed in the method of the invention are degenerate primer sets selected to hybridize with conserved regions of known antibiotic biosynthetic genes, for example Type I and Type II polyketide synthase genes, isopenicillin N synthase genes, and peptide synthetase genes. Each degenerate primer set of the invention includes multiple primer species which hybridize with one DNA strand, and multiple primer species which hybridize with the other DNA strand. All of the primer species within a degenerate primer set which bind to the first strand are the same length, and hybridize with the same target region of the DNA. These primers all have very similar sequences, but have a few bases different in each species to account for the observed variations in the target region. For this reason, they are called degenerate primers.

Similarly, all of the primers within a degenerate primer set which bind to the second strand are the same length, hybridize with the same target region of the DNA, and have very similar sequences with a few bases different in each species to account for the observed variations in the target region.

The degenerate primer sets of the invention are selected to hybridize to highly conserved regions of known antibiotic biosynthesis genes in such a way that they flank a region of several hundred (e.g. 300) or more base pairs such that amplification leads to the selective reproduction of DNA spanning a substantial portion of the antibiotic biosynthesis gene. Selection of primer sets can be made based upon published sequences for classes of antibiotic biosynthesis genes.

For example, for amplification of Type I polyketide synthase genes, we have designed primers based upon the conserved sequences of six beta-ketoacyl carrier protein synthase domains of the erythromycin gene cluster. Donadio et al ., *Science* 252: 675-679 (1991); Donadio and Staver, *Gene* 126: 147-151 (1993). These primers have the sequences

5'-GC(C/G) (A/G)T(G/C) GAC CCG CAG CG CGC-3' [SEQ ID No. 1]

and

5'-GAT (C/G)(G/A)C GTC CGC (G/A)TT (C/G)GT (C/G)CC-3' [SEQ ID No. 2].

The expected size of the PCR product is 1.2 kilobase pairs. Other degenerate primer sets for Type I and Type II polyketide synthetase genes could be determined from sequence information available in Hutchinson and Fujii, *Ann. Rev. Microbiol.* 49: 201-238 (1995).

Type II polyketide synthase gene clusters are characterized by the presence of chain length factor genes which are arranged at the 3'-end of the ketosynthase genes. Primers were designed based on one conserved region near the 3'-end of the ketosynthase gene and one at the middle portion of the chain length factor gene. The sequences of one suitable set of amplification primers are:

5' CT(C/G)AC(G/C)(G/T)(C/G)GG(C/G)CGIAC(C/G)GC(C/G)AC(C/G)CG-3'SEQ ID No. 3

and

5' GTT(C/G)AC(C/G)GCGTAGAACCA(C/G)GCGAA-3' SEQ ID No. 4

The expected size of the PCR product was 0.5 kilobase pairs. An alternative set of degenerate primers has the sequence

5'-TTCGG(C/G)GGITTCCAG(T/A)(C/G)IGC(C/G)ATG SEQ ID No. 5

and

5'-TC(C/G)A(G/T)(C/G)AG(C/G)GC(C/G)AI(C/G)GA(C/G)TCGTAICC SEQ ID No. 6.

These primers were designed based upon consensus sequences for the regions flanking the Ks_{β} (chain length factor) genes. The consensus sequences are available from Hutchinson and Fujii, *supra*.

Primers were designed for beta-lactam biosynthetic genes on the basis of the conserved sequences of a number of isopenicillin N synthase genes as described in Aharanowitz et al., *Ann. Rev. Microbiol.* 46: 461-495 (1992). These primers have the sequences

5'-GG(C/G/T) TC(C/G) GG(C/G) TT(C/T) TTC TAC GC-3' [SEQ ID No. 7]

and

5'-CCT (C/G)GG TCT GG(A/T) A(C/G)A G(C/G)A CG-3' [SEQ ID No. 8].

The expected size of the PCR product is 570 base pairs. Other degenerate primer sets could be determined from sequence information available in Jensen and Demain, "Beta-Lactams" in *Genetics and Biochemistry of Antibiotic Production* (L.C. Vining and C, Studdard, eds.), pp 239-268, Butterworth-Heinemann, Newton, MA (1995).

For isolation of peptide synthetase genes, primers based on two of the conserved core sequences within the functional domains of peptide synthetase genes as described by Turgay and Marahiel, *Peptide Res.* 7: 238-241 (1994) were utilized. These primers had the sequence

5'-ATCTACAC(G/C)TC(G/C)GGCAC(G/C)AC(G/C)GGCAAGCC(G/C)AAGGG-3'

SEQ ID No. 9

and

5'-A(A/T)IGAG(T/G)(C/G)ICCICC(G/C)(A/G)(A/G)(G/C)I(A/C)GAAGAA-3'

SEQ ID No. 10

The expected size of the PCR product is 1.2 kilobase pairs.

PCR amplification can also be used for isolating lichen-derived antibiotic biosynthesis genes and gene fragments. For isolation of Type I polyketide synthase genes from lichens, the primer set used was previously described by Keller et al. in *Molec. Appl. to*

Food Safety Involving Toxic Microorganisms, J.L. Richard, ed., pp. 2630277 (1995), and had the following sequences.

5'-MGIGARGCIYTIGCIATGGAYCCICARCARMG SEQ ID No. 11

and

5'-GGRTCNCCIARYTGIGTICCIGTICCRTGIGC SEQ ID No. 12

The expected size of the PCR product is approximately 0.7 to 0.9 kilobases. Actual products evaluated ranged in size from 637 to 809 nucleotides (not including the 61 nt due to the primers).

Once the primers and the sample are cycled through sufficient thermal cycles to selectively amplify antibiotic biosynthetic DNA in the sample (generally around 25 cycles or more), the amplified DNA is isolated from the amplification mixture. Isolation can be accomplished in a variety of ways. For example, the PCR products can be isolated by electrophoresis on an agarose or polyacrylamide gel, visualized with a stain such as ethidium bromide and then excised from the gel for cloning. Primers modified with an affinity binding moiety such as biotin may also be used during the amplification step, in which case the affinity binding moiety can be used to facilitate the recovery. Thus, in the case of biotinylated primers, the amplified DNA can be recovered from the amplification mixture by coupling the biotin to a streptavidin-coated solid support, for example Dynal streptavidin-coated magnetic beads.

It will be appreciated that the DNA obtained as a result of this isolation will not generally be of a single type because of the degeneracy of the primers and the complexity of the initial sample. Thus, although these steps are sufficient to recover antibiotic biosynthesis genes from soil or lichen, it is preferable to further separate and characterize the individual species of amplified DNA.

This further separation and characterization can be accomplished by inserting the amplified DNA into an expression vector and cloning in a suitable host. The specific combination of vectors and hosts will be understood by persons skilled in the art, although bacterial expression vectors and bacterial hosts are generally preferred. Individual clones are then picked and the sequence of the cloned plasmid determined. While random selection has been employed successfully, selection of antibiotic biosynthesis gene-containing clones

can be facilitated by screening using hybridization with DNA probes based on conserved sequences or by overlay of bacterial clones with an antibiotic-sensitive test strain.

Once the sequence of the cloned DNA is determined, it can be screened against existing libraries of nucleotide and protein sequences for confirmation as an antibiotic biosynthetic gene or gene fragment. Amplified DNA so-identified can be used in several ways. First, the amplified DNA, or distinctive portions thereof, can be used to as probes to screen libraries constructed from humic-derived or lichen DNA to facilitate the identification and isolation of full length antibiotic biosynthetic genes. Once isolated, these genes can be expressed in readily cultivated surrogate hosts, such as a *Streptomyces* species for soil-derived genes or an *Aspergillus* species for lichen-derived genes. General procedures for such expression are known
in the art, for example from Fujii et al., *Molec. Gen. Genet.* 253: 1010 (1996) and Bedford et al., *J. Bacteriol.* 177: 4544-4548 (1995), which are incorporated herein by reference. Second, amplified DNA which is different from previously known DNA can be used to generate hybrid antibiotic biosynthesis genes using the procedures described by McDaniel et al, *Nature* 375: 549-554 (1995); Stachelhaus et al., *Science* 269: 69-72 (1995); and Stachelhaus et al, *Biochem, Pharmacol.* 52: 177-186 (1996). In these procedures, the novel DNA sequences isolated using the method of the invention are spliced into a known antibiotic gene to provide an expressible sequence encoding a complete gene product.

Using the method of the invention, a number of unique nucleotide sequences have been identified and characterized. The sequences and the biosynthetic polypeptides/proteins for which they encode, given by sequence ID Nos. 13 to 80, are a further aspect of the present invention.

EXAMPLE 1

Total DNA was extracted from soil samples by a direct lysis procedure as described by Barns et al. (1994). The high molecular weight DNA (>20 kb) in the extract was separated on a Sephadex G200 column (Pharmacia, Uppsala, Sweden) as described by Tsai and Olson, *Appl. Environ. Microbiol.* 58: 2292-2295 (1992).

The DNA extract (10-50 ng template DNA) was added to an amplification mixture (total volume 100 µl) containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM

$MgCl_2$, 200 μM of each deoxynucleotide triphosphate, 25 pmol of each Type I polyketide primer (Seq ID Nos 1 and 2) and 5.0 units of Taq polymerase (BRL Life Technologies, Gaithersburg, MD). The mixture was then thermally cycled for 30 cycles in a MJ Research PTC-100 thermocycler using the following program:

denaturation	93°C	60 seconds
annealing	60°C	30 seconds
extension	72°C	90 seconds

The PCR products were then electrophoresed in 1% agarose gels and stained with ethidium bromide to visualize the DNA bands. Bands containing PCR product of the expected size were excised from the gel and purified using a Qiaex Gel Extraction kit (Qiagen GmBH). The purified DNA was ligated to pCRII (Invitrogen) to generate a clone library using *E. coli* INVαF competent cells. 18 clones were chosen at random from the library and sequenced using a Taq Dye Terminator Cycle Sequencing Kit and an Applied Biosystem DNA sequencer model 373. The sequencing primers used included the universal M13 (-20) forward primer, the M13 reverse primer and primers designed from the sequence data obtained. DNA sequences were translated into partial amino acid sequences using a software package from Geneworks (Intelligenetics, Inc.) with further manual adjustments and sent to the NCBI database by e-mail at blast@ncbi.nlm.nih.gov for comparison against protein databases. Altschul et al., "Basic Local Alignment Tool", *J. Mol. Biol.* 215: 403-410 (1990).

Blast analysis of the 18 clones pointed to 12 unique sequences that were not identical to each other or to published sequences. Seq. ID No. 13 shows the complete DNA sequence of a representative unique clone (Clone ksfs). Seq. ID No. 14 shows the translated amino acid sequence of this clone. The greatest homology as determined by a Blast analysis is indicated to be Type I polyketide synthases. Similar results were obtained on the Blast search of the other 11 unique clones based upon partial sequences which were determined.

EXAMPLE 2

The experiment of Example 1 was repeated using isopenicillin N synthase gene primers (Seq ID Nos. 7 and 8). The thermal cycling program was changed to include 60 second extension periods at 72°C, but otherwise the experimental conditions were the same. Twelve clones were picked at random and yielded one unique sequence that was not identical

to published sequences. The complete sequence of this clone (Clone ipnsfs) is shown in Seq. ID. No. 15 and the translated amino acid sequence in Seq. ID No. 16. The BLAST search indicated greatest homology for this sequence with isopenicillin N synthases.

EXAMPLE 3

The experiment of Example 1 was repeated using peptide synthetase primers (Seq. ID Nos 9 and 10). The amplification mixture was changed to a 50 ul volume containing 10 to 50 ng of template DNA, 20 mM $(NH_4)_2SO_4$, 74 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 0.01% Tween 20, 200 µM of each deoxynucleotide triphosphate, 25 pmol of each primer, 0.25 % skim milk and 0.4 units of Ultra Therm DNA Polymerase (Bio/Can Scientific, Mississauga, Ontario). The mixture was thermocycled for 30 cycles using the following program:

denaturation	95°C	60 seconds
annealing	52°C	60 seconds
extension	72°C	120 seconds.

Thirty clones containing a 1.2 kb insert have been partially sequenced. The BLAST analysis of the 30 clones pointed to 28 unique sequences that were not identical to each other or to published sequences. Varying degrees of homology to known peptide synthase genes were seen. Seq. ID No. 17 shows the complete DNA sequence of representative clone (ps32). Seq. ID No. 18 shows the translated amino acid sequence of this clone. Based on a Blast search of these sequences, the greatest homology is to a peptide synthase gene such as the pristinamycin synthase gene from *Streptomyces pristinaespiralis* and *Bacillus* sp. peptide synthetase genes such as gramicidin S synthase and surfactin synthetase. Stachelhaus and Marahiel, *FEMS Micro. Letters* 125: 3-14 (1995); Turgay et al., *Mol. Micro* 6: 529-546 (1992).

Sequence ID Nos. 81 to 94 show an additional 7 unique sequences (nucleic acid and translated amino acid sequences) of 1.2 kb PCR products amplified from soil DNA samples using these primers. These sequences have been named ps 2, ps 3, ps 7, ps 10, ps 24, ps 25 and ps 30. The sequences are unique in that they are all different from each other and from ps 32,

and while they show greatest homology to peptide synthetase sequences in the databases searched by BLAST analysis, they do not match any known sequence. Within each, the conserved motifs (TGD, KIRGXRIEL, NGK) common to peptide synthetase domains as described by Turgay and Marahiel (1994) can be identified. Descriptive information of the clones follows:

Clone ps 2, 1204 bp, with conserved motifs SGD, KIRGFRIEL, NGK, 67% G + C

Clone ps 3, 1178 bp, with conserved motifs TGD, KIRGSRIEL, NGK, 59 % G + C

Clone ps 7, 1222 bp with conserved motifs TGD, KIRGYRIEL, NGK, 55.5 % G + C

Clone ps 10, 1171 bp with conserved motifs TGD, KIRGHRIEL, NLK, 63% G + C

Clone ps 24, 1190 bp with conserved motifs TGD, KIRGHRIAM, NQK, 56 % G + C

Clone ps 25, 1178 bp with conserved motifs TGD, KLRGYRIEL, NDK 68 % G + C

Clone ps 30, 1200 bp with conserved motifs TGD, KVRGFRIEP, NGK, 64.5 % G + C

Clone ps 32, 1172 bp with conserved motifs TGD, KIRGFRIEL, SGK, 67 % G + C

EXAMPLE 4

The experiment of example 1 was repeated using the Type II polyketide synthase primers given by Seq. ID. Nos. 3 and 4. PCR amplification was carried out in a total volume of 50 ul containing 50 ng of soil DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl , 2 mM $MgCl_2$, 200 uM of each deoxynucleotide triphosphate, 25 pmol of each primer and 5.0 units of *Taq* polymerase (BRL Life Technologies, Gaithersburg, MD). The thermal cycling conditions included denaturations at 94°C for 60 seconds, annealing at 58°C for 30 seconds and extensions at 72°C for seconds, repeated for a total of 30 cycles.

PCR amplification yielded products of the expected size of 0.5 kilobase pairs. Sequencing of 18 randomly selected clones revealed the presence of 5 unique sequence that

were not identical to each other or to published sequences. Seq. ID No. 19 shows the complete DNA sequence of a representative clone (clone clf). The translated amino acid sequence of this clone is shown in Seq. ID. No. 20. In a BLAST search of this DNA sequence against the protein database, the greatest homology is indicated to chain length factor genes of the Type II polyketide synthases.

Example 5

The experiment of Example 1 was repeated using the Type I polyketide synthase primers designed for fungal sequences. (Seq. ID. Nos. 11 and 12) PCR amplifications were carried out with lichen DNA samples from a variety of lichen species representing 11 genera prepared as described in Miao et al. (1991), *supra.*

PCR amplifications were carried out in a total volume of 50 ul containing approximately 10 ng of lichen DNA and 1 unit of *Taq* polymerase in a reaction as per Example 4. The cycling protocol was 30 cycles of denaturation at 95°C for 60 seconds, annealing at 57°C for 2 minutes and extensions at 72°C for 2 minutes.

Forty seven clones with inserts of the expected size have been partially sequenced. The sequences all show homology to Type I fungal polyketide synthase genes but are all distinct from each other and from known sequences. Seq. ID. No. 21 shows the complete DNA sequence of a 637 base pair product amplified from DNA extracted from the lichen *Xanthoparmelia cumberlandia* (clone Xa.cum.6A). The translated amino acid sequence is shown in Seq. ID. No. 22. The greatest homology as determined by Blast analysis is indicated to fungal Type I polyketide synthase genes. Sequence ID Nos. 29 and 30 show the DNA sequence and conceptual amino acid sequence, respectively, for a further clone Xa.cum.6H isolated in this experiment. Sequences of DNA and the corresponding amino acid sequences for seven other lichen samples, *Leptogium corniculatum* (Seq. ID Nos. 31-42), *Parmelia sulcata* (Seq. ID Nos. 43-50); *Peltigera neopolydactyla* (Seq. ID Nos. 51-60); *Pseudocyphellaria anthrapsis* (Seq. ID Nos. 61-62); *Siphula ceratities* (Seq. ID. Nos. 63-66); *Thamnolia vermicularis* (Seq. ID Nos. 67-68); and *Usnea florida* (Seq. ID Nos. 69-80). Each of these sequences showed homology by Blast analysis to fungal Type I polyketide synthase.

EXAMPLE 6

The experiment of Example 5 was repeated on DNA from the lichen *Solorina crocea* using the degenerate peptide synthetase primers of Example 3. Freshly collected lichen (approximately 1.2 g) was washed in running tap water to remove conspicuous soil and field detritis, and then further cleaned under a dissecting microscope. The cleaned sample was then gently shaken in a 50 ml tube containing about 40 ml of 0.2% SDS for at least 30 minutes and rinsed thoroughly with water. Excess surface water was blotted from the washed, hydrated lichen, and the sample was frozen at -80°C for at least 15 minutes then vacuum dried at room temperature for 4 hours. The lichen was ground in liquid nitrogen using a mortar and pestle to produce a lichen powder for use in preparing DNA extracts.

To prepare the DNA extracts, 0.28g of lichen powder was placed into 18 2-ml microfuge tubes, and each aliquot was mixed with 1.25 ml isolation buffer (150 mM EDTA, 50 mM Tris pH 8, 1% sodium lauroyl sarcosine) and extracted for 1 hour at 62°C. The samples were centrifuged for three minutes to pellet cellular debris and a cloudy supernatant was decanted into new microfuge tubes. Each sample of the supernate was mixed with 750 µl 7.5 M ammonium acetate, incubated on ice for 30 minutes and centrifuged for five minutes at 16,000 X g to precipitate proteins. The supernatant fluid was saved in new microfuge tubes and nucleic acids were precipitated with 0.6 volumes of isopropanol overnight at 4°C. Samples were centrifuged for five minutes at 16,000 X g to pellet nucleic acids. The pellets were dissolved in TE containing RNAse (18 µg total) at 50°C for 45 minutes. The solutions were then extracted with an equal volume of TE saturated phenol:chloroform (1:1), and again with chloroform. DNA in the aqueous phase was precipitated with 0.1 M sodium acetate and two volumes of ethanol at -20°C for 2 hours, and then pelleted by centrifugation for five minutes at 16,000 X g. The DNA pellet was washed with 75% ethanol, vacuum dried at room temperature for 3 minutes and then dissolved in TE. The final amount of DNA recovered was approximately 70µg according to fluorometric measurement.

Two clones containing the expected 1.2 kb insert were sequenced and found to contain the same sequence shown in Seq. ID. No. 23. Seq. ID. No. 24 shows the translated amino acid sequence. The sequence is distinct, with greatest homology as determined by Blast analysis to the peptide synthase module of the cyanobacterium *Microcyctis aeruginosa*.

EXAMPLE 7

The experiment of example 4 was repeated using the Type II polyketide synthase primers given by Seq. ID. Nos. 5 and 6. Three starting samples were used for recovery of Type II polyketide synthase genes: two uncharacterized strains of *Streptomyces* (strains WEC 68A and WEC 71B) which had been shown to contain Type II polyketide synthase genes, and a soil sample obtained from a forest area near Vancouver, British Columbia. The soil sample was prepared using the basic protocol from Holben et al, *Appl. Environ. Microbiol.* 54: 703- 711 (1988) with variations in parameters such as mix time to adjust for the individual characteristics of the soil samples.

Streptomyces genomic DNA preparations suitable for PCR amplification were prepared from the mycelia harvested from a 50 ml culture in tryptic soy broth (Difco) which had been grown for 3 days at 300 C. The mycelia were collected by centrifugation at 2500 x g for 10 minutes, the pellets were washed in 10% v/v glycerol and the washed pellets were frozen at -200C. The size of the pellets will vary with different strains; for extraction, 1 g samples were suspended in 5 ml TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) in a 50 ml screw cap Oakridge tube and lysozyme (to 10 mg/ml) and RNase (to 40 ug/g) were added. Following incubation at 300C for 45 min. a drop of each suspension was transferred to a microscope slide, one drop of 10% SDS was added and the suspension was checked for complete clearing and increased viscosity, indicating lysis. Most strains lyse with this incubation time, but incubation in lysozyme may be continued if necessary. (For strains which are very resistant to lysis, small amounts of DNA suitable for PCR amplification may often be prepared on a FastPrep™ instrument as described below.) Following confirmation of sufficient incubation time in lysozyme, 1.2 ml of 0.5 M EDTA, pH 8.0 was added to the suspension and mixed gently then 0.13 ml of 10 mg/ml Proteinase K (Gibco/BRL) solution was added and incubated for 5 min. at 300 C. 0.7 ml of 10% SDS was added, mixed gently by tilting, then incubated again at 300 C for 2 hours. Following lysis, three successive phenol/chloroform extractions were performed by adding a volume equivalent to the aqueous phase each time of a 1:1 mixture of ultrapure Tris buffer saturated phenol (Gibco/BRL) and chloroform. The aqueous phase was recovered each time following centrifugation at 2500 x g for 10 min. in a shortened (i.e.wide bore) Pasteur pipet to minimize shearing; DNA was precipitated from the final aqueous phase with the addition of 0.1 volume of 3M Na acetate,

pH 4.8 and 1 volume of isopropanol at room temperature. DNA was spooled from the solution onto a sealed Pasteur pipet, rinsed in ice cold 70% ethanol and solubilized in 0.5 ml TE buffer overnight at room temperature. DNA yields (as determined spectrophotometrically) typically range from 1 to 3 mg from 1 g of mycelia.

An alternative method for the preparation of small amounts of Streptomyces DNA suitable for PCR amplification has been found to be useful for strains resistant to lysis or when a faster method is desirable. This method makes use of the FastPrep™ instrument (Savant) and the methods and kit supplied by BIO 101 (Bio/Can Scientific, Mississauga, Canada). A 2 ml aliquot from a 20 ml, 3 day culture in tryptic soy broth is pelleted in a 2 ml microfuge tube and the size of the mycelial pellet is estimated. "Small" pellets are resuspended in 100 ul of sterile distilled water; larger pellets are resuspended in 200-300 ul of water. 200 ul of suspension is transferred to a homogenization tube from the kit . Following the manufacturer's protocol for the preparation of DNA from medium hard tissue, the large bead is added to this tube (which already contains a small bead) and 1 ml of solution CLS-TC from the kit is added and the samples are processed in the instrument for 10 seconds at speed setting 4.5. Samples are then spun 15 min. at 10,000 x g at 40C and 600 ul of the supernatant is transferred to a clean microfuge tube, 400 ul of Binding Matrix is added and mixed gently, then the sample is spun for 1 min. as above. The supernatant is discarded while the pellet is resuspended in 500 ul SEWS-M and transferred to a SPIN™ Filter unit. This is spun for 1 minute, the contents of the catch tube are discarded and the unit is spun again to dry. The filter unit is transferred to a new microfuge tube and DNA is eluted from the matrix in 100 ul DES which is left on the filter for 2-3 min. at room temperature. Eluted DNA is collected by spinning once again and this DNA is now ready to use in PCR amplifications. Due to components of the final solution, DNA prepared by this method is difficult to quantify. Typically 1 ul or 1/10 ul of this eluate is suitable as a template for PCR; larger quantities may be inhibitory to the PCR polymerase.

PCR amplification was carried out in a total volume of 50 ul containing 50 ng of DNA, 5 % DMSO, 1.25 mM $MgCl_2$, 200 uM of each deoxynucleotide triphosphate, 0.5 ug of each primer and 5.0 units of *Taq* polymerase (BRL Life Technologies, Gaithersburg, MD). The thermal cycling started with a 'touch-down' sequence, lowering the annealing temperature from 65°C to 58°C over the course of 8 cycles. The temperature of the annealing step

was then maintained at 58°C for a further 35 cycles. The overall cycle used was: denaturation at 94°C for 45 seconds, annealing at 65°C to 58°C for 1 minute and extension at 72°C for 2 minutes. The size of the amplified fragments was expected to be approximately 1.5 kb.

Amplification of the two *Streptomyces* strains produced DNA fragments of the expected size (1482 bp and 1538 bp). Open reading frame analysis of the two sequences revealed the presence of a set of three ORFs each, corresponding to the 3'-ends of the putative Ks_{α}-subunit genes (50 to 60 bp), possible full-length Ks_{β} genes (approx. 1.2 kb) and the first halves of potential ACP genes (approx 100 bp). In each sequence, the first and second ORFs were linked by a stop codon overlap typical of $Ks_{\alpha,\beta}$ gene pair junctions and a possible indication of tight coexpression through translational coupling. The two Ks_{β} genes were separated from the downstream ACP genes by a short spacer, again consistent with the expected gene organization.

Two clones were selected from among clones created using the soil DNA as a source which were found to produce 1.5 kb inserts. These inserts were sequenced and found to exhibit similarity to known KS_{β} genes with three ORFs as described above. The translated amino acid sequences of the four genes are shown in Sequence ID Nos 25 to 28.

The four putative KS_{β} genes had G+C content over 70% which is typical for the coding regions of Actinomycete genes. Results of data base searches established that the deduced products of all four ORFs were similar to known KS_{β} gene products from Type II polyketide synthases but they did not match any known sequences.

EXAMPLE 8

DNA can be extracted from large volumes of soil in accordance with the following procedure. Place dry soil into a sterile blender with 0.2% sodium pyrophosphate (100 ml/100 grams of soil). The pH of the sodium pyrophosphate solution should be about 10, although some variation to account for the characteristics of the soil may be appropriate. The mixture is blended for 30 seconds, decanted into centrifuges bottles and then centrifuged for 15 minutes at 100 X g at 4°C. The supernatant is decanted, filtered two times through cheese cloth and saved. The pelleted soil is extracted an additional two times using the same procedure.

After the extractions, the pooled supernatants are centrifuged for 15 minutes at 10,500 X g and the pellets are collected. The pellet may be incubated for 6 hours at 55°C in pre-germination medium (0.5% w/v yeast extract (Difco), 0.5% w/v casamino acids (Difco) with 0,005 M $CaCl_2$ and 0.025 M TES, pH 8.0 (added separately from sterile stock after autoclaving other components)) and then repelleted, or it may be used directly. In either case, the pellet (approximately 30-200 mg) is mixed with 5 ml 1X TE (pH 8.0), 500 µl 0.5M EDTA (pH 8.0) and 500 µl - 20 mg/ml lysozyme in 1X TE (pH 8.0) and incubated for 30 minutes at 37°C. 500 µl of 20% SDS and 100 µl - 1% proteinase K in TE and 1% SDS are then added and the mixture is vortexed gently before incubating for 60 minutes at 55°C or overnight at 37°C.

The incubated mixture is combined with 10 ml 20% polyvinylpyrrolidone (avg. MW=40,000) and incubated for 10 minutes at 70°C. One-half volume of 7.5 M ammonium acetate (stored at -20°C) is then added, the resulting mixture is placed for 10 minutes on a low speed shaker, and then centrifuged for 20 minutes at 18,5000 X g. The supernatant is combined with 1 volume of isopropanol and incubated for 30 minutes at -20°C before centrifuging for 20 minutes at 18,500 X g. The pellet from this centrifugation is washed in 70% ethanol, and centrifuged for 10 minutes at 18,500 X g. The pellet from this final centrifugation is collected and air dried.

EXAMPLE 9

To extract DNA from small amounts of soil the following procedure can be used. Combine soil (approx 1 g) with 1 ml distilled water, vortex to suspend and pellet at 19,000 X g for 5 minutes. After removing the supernatant, freeze/thaw the samples twice by either of the following techniques (a) -20°C freezer, 30 minutes, followed by 50-60°C water bath (2 minutes), repeated 2 times; or (b) quick freeze in EtOH-dry ice bath (dip in until frozen, approx one minute) followed by 60°C water bath (2 minutes), repeated 2 times. The pellets are then suspended in 350 µl TE buffer (pH 8.0), 50 µl 0.5 M EDTA and 50 µl-20 mg/ml lysozyme in TE buffer, vortexed and incubated at 37°C for 30 minutes in a water bath. 50 µl of 20% SDS and 10 µl 1% Proteinase K/ 1% SDS in TE buffer is added, vortexed, and incubated for one hour at 55°C or overnight at 37°C. One-tenth volume of 20% polyvinylpyrrolidone (avg. MW=40,000) is then added and incubated at 70°C for 10 minutes.

One-half volume of 7.5 M ammonium acetate (stored at -20°C) is added, the tubes are shaken at low speed for ten minutes and then centrifuged at 19,000 X g for 20 minutes. The supernatant is collected using pipets with cut tips to avoid shearing DNA, combined with one volume of isopropanol, mixed gently, and stored at -20°C for 30 minutes or 4°C overnight. The DNA is then collected as a pellet by centrifugation at 19,000 X g for 10 minutes. The resulting pellet is washed with 0.5 ml of 70% ethanol (stored at -20°C) and then air or vacuum dried. The dried DNA is then dissolved in 50-150 ul of TE buffer, incubated at 4°C for one hour and then heated to 60°C for 10 minutes to facilitate dissolving DNA. The resulting solutions are stored at -20°C until use.

SEQUENCE LISTING

(1) GENERAL INFORMATION:
(i) APPLICANT: Terragen Diversity Inc.
(ii) TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR BIOACTIVE MOLECULES
(iii) NUMBER OF SEQUENCES: 94
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Deeth Williams Wall
(B) STREET: National Bank Building, 150 York Street, Suite 400
(C) CITY: Toronto
(D) STATE: Ontario
(E) COUNTRY: Canada
(F) ZIP: M5H 3S5
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
(B) COMPUTER: Dell (IBM Compatible)
(C) OPERATING SYSTEM: Windows 95
(D) SOFTWARE: Word 97
(vi) CURRENT APPLICATION DATA :
(A) APPLICATION NUMBER: Not yet assigned
(B) FILING DATE: May 21, 1998
(C) CLASSIFICATION:
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/861,774
(B) FILING DATE: May 22, 1997
(viii) ATTORNEY/AGENT INFORMATION :
(A) NAME: Eileen McMahon
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER: 1694/0005
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 416-941-9440
(B) TELEFAX: 416-941-9443
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
GCSRTSGACC CGCAGCGCGC 20

(2) INFORMATION FOR SEQ ID NO:2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
GATSRCGTCC GCRTTSGTSC C 21

(2) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
CTSACSKSGG SCGNACSGCS ACSCG 25

(2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
GTTSACSGCG TAGAACCASG CGAA 25

(2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
TTCGGSGGNT TCCAGWSNGC SATG 24

(2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA

(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
TCSAKSAGSG CSANSGASTC GTANCC 26

(2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
GGBTCSGGST TYTTCTACGC 20

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
CCTSGGTCTG GWASAGSACG 20

(2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
ATCTACACST CSGGCACSAC SGGCAAGCCS AAGGG 35

(2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AWNGAGKSNC CICCSRRSNM GAAGAA 26

(2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
MGIGARGCIY TIGCIATGGA YCCICARCAR MG 32

(2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
GGRTCNCCIA RYTGIGTICC IGTICCRTGI GC 32

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1206
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GCGGTGGACC CGCAGCAGCG CCTCATGCTG GAGCTGGCCT GGTCCGCGCT     50
GGAAAGCGCA GGTCATCCGC CCTCGATATT CCCCGGCCTG ATCGGGGTCT    100
ATGTCGGCAT GAACTGGAAT CGCTATCGCG CGAATTGCAT TTCTGCACAC    150
CCTGATGTGG TGGAGCGATT CGGTGAATTG AACACAGCGC TCGCCAACGA    200
ATACGACTTT CTTGCTACCC GAATCTCCTA CAAGCTCAAT CTGCGCGGTC    250
CCAGCGTCAC TATCAGCACC GCTTGTTCGA CTTCCCTGGT TGCCATTGCT    300
CAGGCTTCGC AGGCGTTGCT CAACTATGAA TGCGACATTG CTTTGGCTGG    350
GGTTGCCTCC ATAACCGTGC CTGTCAATGC AGGCTACCTC TACCAAGAAA    400
GGTGGCATGC TTTCACCGAA GGGCATTGTC CTACATTCGA TGCCCCAGCA    450
CGGGACCACT TCAATGATGC CCCCTGTCTC CTTTTTGCGG GCCTGGAAAA    500
CCCATCCAGG AGGGGGGGGG GGGCCCTCAT ACCCGGCCTT TCAAGCGGGA    550
ACCTCTCACA GGAAGCGGAT GTTTCAGCCG AAGGGATGTT GAACATTGAC    600
GCCGGCAGCA CGGGGGACAA GTTCAGGGAT GGGCGCGCTT TTGTTGTATG    650
GGGGGGGCCT GGAAGAAGCA TTCAAGGGAC GGTGATCAAA CTTAACCCCT    700
TCATTGGCGG GTTTGCCGCG GAACAAGGAC GGGTTCGGAC AAGGCGAGTT    750
TACCGGCGCC CAGGCGTCAA TGGTCAGGGC GGAGTTCATT TCGCTTTGGC    800
```

GGTGGAGTTT GCGGGATATT CGAATCCCGC AAGCATCGGG ATTTCATTCG 850
AAAACCCACG GGCACGGGCG ACGCCATTGG GCGATCCGAT AGAAGTGGCC 900
GCGCTAAAGA TGGTTTTTCG CCGACGCTCG TTCCAGAGGC GCCGTTGCGC 950
CCTTGGATCG GTCAAGAGTT GTGTCGGACA CCTGGTTCAC GCCGCCGGCG 1000
TGACCGGATT TATCAAGGCT GTCTTGTCGG TCTACCACGG CAAGATCGCA 1050
CCGACACTGT TTTTCGAGAA AGCAAATCCG AGGCTCGGGC TGGAAGACAG 1100
TCCTTTCTAT GTCAATGCCG GACTCGAGAA GTGGACGGCC GCCGAGCAGC 1150
CACGCCGCGC GGGGGTCAGT GCTTTCGGGG TCGGTGGCAC CAATGCGCAC 1200
GCGATC 1206

(2) INFORMATION FOR SEQ ID NO:14:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ala Val Asp Pro Gln Gln Arg Leu Met Leu Glu Leu Ala Trp Ser
5 10 15

Ala Leu Glu Ser Ala Gly His Pro Pro Ser Ile Phe Pro Gly Leu
20 25 30

Ile Gly Val Tyr Val Gly Met Asn Trp Asn Arg Tyr Arg Ala Asn
35 40 45

Cys Ile Ser Ala His Pro Asp Val Val Glu Arg Phe Gly Glu Leu
50 55 60

Asn Thr Ala Leu Ala Asn Glu Tyr Asp Phe Leu Ala Thr Arg Ile
65 70 75

Ser Tyr Lys Leu Asn Leu Arg Gly Pro Ser Val Thr Ile Ser Thr
80 85 90

Ala Cys Ser Thr Ser Leu Val Ala Ile Ala Gln Ala Ser Gln Ala
95 100 105

Leu Leu Asn Tyr Glu Cys Asp Ile Ala Leu Ala Gly Val Ala Ser
110 115 120

Ile Thr Val Pro Val Asn Ala Gly Tyr Leu Tyr Gln Glu Arg Trp
125 130 135

His Ala Phe Thr Glu Gly His Cys Pro Thr Phe Asp Ala Pro Ala
140 145 150

Arg Asp His Phe Asn Asp Ala Pro Cys Leu Leu Phe Ala Gly Leu
155 160 165

Glu Asn Pro Ser Arg Arg Gly Gly Gly Ala Leu Ile Pro Gly Leu

170 175 180

Ser Ser Gly Asn Leu Ser Gln Glu Ala Asp Val Ser Ala Glu Gly
185 190 195

Met Leu Asn Ile Asp Ala Gly Ser Thr Gly Asp Lys Phe Arg Asp
200 205 210

Gly Arg Ala Phe Val Val Trp Gly Gly Pro Gly Arg Ser Ile Gln
215 220 225

Gly Thr Val Ile Lys Leu Asn Pro Phe Ile Gly Gly Phe Ala Ala
230 235 240

Glu Gln Gly Arg Val Arg Thr Arg Arg Val Tyr Arg Arg Pro Gly
245 250 255

Val Asn Gly Gln Gly Gly Val His Phe Ala Leu Ala Val Glu Phe
260 265 270

Ala Gly Tyr Ser Asn Pro Ala Ser Ile Gly Ile Ser Phe Glu Asn
275 280 285

Pro Arg Ala Arg Ala Thr Pro Leu Gly Asp Pro Ile Glu Val Ala
290 295 300

Ala Leu Lys Met Val Phe Arg Arg Arg Ser Phe Gln Arg Arg Arg
305 310 315

Cys Ala Leu Gly Ser Val Lys Ser Cys Val Gly His Leu Val His
320 325 330

Ala Ala Gly Val Thr Gly Phe Ile Lys Ala Val Leu Ser Val Tyr
335 340 345

His Gly Lys Ile Ala Pro Thr Leu Phe Phe Glu Lys Ala Asn Pro
350 355 360

Arg Leu Gly Leu Glu Asp Ser Pro Phe Tyr Val Asn Ala Gly Leu
365 370 375

Glu Lys Trp Thr Ala Ala Glu Gln Pro Arg Arg Ala Gly Val Ser
380 385 390

Ala Phe Gly Val Gly Gly Thr Asn Ala His Ala Ile
395 400

(2) INFORMATION FOR SEQ ID NO:15
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 565
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
GGCTCCGGGT TTTTCTACGC GTCCAACCAC GGGATCGACG TCACGCGGGT    50
GCGCGACGAG GTGAACAAGT TCCACGCCGA GATGACGCCC GGGGAGAAGT   100
TCGAGCTGGC CATCAACGCC TACAACGACG CGAATCCGCA TACCCGCAAC   150
GGGTATTACA TGGCCGTCGA AGGCAAGAAG GCCGTCGAGT CCTTCTGCTA   200
CCTCAACCCG GCCTTCACCC CCGAGCACCC GATGATCGAG GCGGGCGCGG   250
CGGGGCACGA GGTGAACAAC TGGCCGGACG AGGCTCGCCA CCCCGGCTTC   300
CGTGAGTACG GGGGAGCAGT ACTTCGAAGA GGATCCTCCG ACCTGTCACT   350
GGTGCTGCTG CGTGGGTACG CGCTGGCCCT GGGCAAGGAC GAGAACTACT   400
TCGACGACTA CGTCAAGCAC TCCGACACGC TCTCGGCCGT CTCGCTGATC   450
CGTTACCCGT ACCTGGAGAA CTACCCGCCG GTGAAGACCG GTCCGGACGG   500
CGAGAAGCTC AGCTTCGAGG ATCACTTCGA CGTCTCGCTG ATCACCGTGC   550
TCTTCCAGAC CCAGG                                        565
```

(2) INFORMATION FOR SEQ ID NO:16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 188
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Gly Ser Gly Phe Phe Tyr Ala Ser Asn His Gly Ile Asp Val Thr
  5                  10                  15

Arg Val Arg Asp Glu Val Asn Lys Phe His Ala Glu Met Thr Pro
 20                  25                  30

Gly Glu Lys Phe Glu Leu Ala Ile Asn Ala Tyr Asn Asp Ala Asn
 35                  40                  45

Pro His Thr Arg Asn Gly Tyr Tyr Met Ala Val Glu Gly Lys Lys
 50                  55                  60

Ala Val Glu Ser Phe Cys Tyr Leu Asn Pro Ala Phe Thr Pro Glu
 65                  70                  75

His Pro Met Ile Glu Ala Gly Ala Ala Gly His Glu Val Asn Asn
 80                  85                  90

Trp Pro Asp Glu Ala Arg His Pro Gly Phe Arg Glu Tyr Gly Gly
 95                 100                 105

Ala Val Leu Arg Arg Gly Ser Ser Asp Leu Ser Leu Val Leu Leu
110                 115                 120

Arg Gly Tyr Ala Leu Ala Leu Gly Lys Asp Glu Asn Tyr Phe Asp
125                 130                 135

Asp Tyr Val Lys His Ser Asp Thr Leu Ser Ala Val Ser Leu Ile
```

```
                140                 145                 150

Arg Tyr Pro Tyr Leu Glu Asn Tyr Pro Pro Val Lys Thr Gly Pro
                155                 160                 165

Asp Gly Glu Lys Leu Ser Phe Glu Asp His Phe Asp Val Ser Leu
                170                 175                 180

Ile Thr Val Leu Phe Gln Thr Gln
                185
```

(2) INFORMATION FOR SEQ ID NO:17:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1172
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
AAGGAGGGGC CGCCCGGGGC GAAGAAGCTG TCCGTCCGAC TGACACGTTC    50
CACTCCGAGG AGCCCGGACC AGATGCGCGC CAGCTTTACC TCGACCGGCG   100
TAGATGGCGG GTCGTAGTCA GTGCGATCCG ATGAGTCATC TGGAGGTGCA   150
GGCAGCACCT TCAGATCGAT CTTGCCGCTC GCCATGCGCG GCATCTCGCG   200
GAGCTCGACG AATGCAGCCG GAATCATGTA CTCGGGCAAC CGCGTGCGAA   250
GATGATCGCG CAGCTCGGAC GCGGCGACCG AGGCGAGCCG AGGCGACCAG   300
TACGCAACGA GACGCTTGTC GCCGGCCCGC TCCTGCCGCG CCAGGACGAC   350
GGCCGTCTCG ACACCGGGGT GATCGGCCAG CGCCGCCTCG ATCTCACCGA   400
GCTCGATGCG GAAGCCGCGG ATCTTGACCT GATGATCCGC GCGCCCGATG   450
AAGTCGAGGT TGCCGTCCGG AAGCCAGCGC ACCAGGTCGC CGGTCCGGTA   500
CAGCCGCGAG CCAGGTGCAC CGAATGGATC GGGTACGAAC CGCGCTCCGG   550
TGAGGGCGGC ATCATCGACA TAGCCGCGCG CGAGGTTCTC GCCACCGATG   600
TACAGCTCGC CGATCACGCG CGCCGGAACG GGCTCGAGTG CGCTATCGAG   650
CACGTAGACC TGAACGTTGT CGAGCGGACG GCCGATCGAC GGCAGCTCGG   700
ACCCGTGTTC GGACGCGGGC GACACGATCG CCCACGTCGT ATCGACCGCG   750
TTCTCCGTCG GGCCGTACTC GTTGAGCATG CGGTAGTGCG CATCGCGCGG   800
TGGACGCCGC GTGAGTCGAT CACCGCCCGT ACGCAGCACG CGCAACGAGC   850
GTGGAAAGTC GCCAGCCGCG AGCAACGCGT CGAGTAGCCG GCCTGGAAGA   900
TCGGAGATCG TGATCCCCCA TCGCGTCAGG TTCTCGAGCA GGCGCGGCGG   950
ATCGAGGCGG AGCTCGTTGT CCACCAGATG AAGCCGGGCG CCCGTCGCCA  1000
GCGTGGACCA CAGCTCGAGC GCCGCGGCAT CGAACGACAT CGAGTAGATC  1050
TGCGTCACGC GGTCGTCGGC ACTGATCTCG ACGGCACGCT GGTTCCACGC  1100
GATCAAATTT CTCAGTGCAC GGTGCGGCAC GGCGACGCCC TTCGGCTTGC  1150
CCGTCGTGCC CGACGTGTAG AT                                 1172
```

(2) INFORMATION FOR SEQ ID NO:18:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein

(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Val Ala Val
5 10 15

Pro His Arg Ala Leu Arg Asn Leu Ile Ala Trp Asn Gln Arg Ala
20 25 30

Val Glu Ile Ser Ala Asp Asp Arg Val Thr Gln Ile Tyr Ser Met
35 40 45

Ser Phe Asp Ala Ala Ala Leu Glu Leu Trp Ser Thr Leu Ala Thr
50 55 60

Gly Ala Arg Leu His Leu Val Asp Asn Glu Leu Arg Leu Asp Pro
65 70 75

Pro Arg Leu Leu Glu Asn Leu Thr Arg Trp Gly Ile Thr Ile Ser
80 85 90

Asp Leu Pro Gly Arg Leu Leu Asp Ala Leu Leu Ala Ala Gly Asp
95 100 105

Phe Pro Arg Ser Leu Arg Val Leu Arg Thr Gly Gly Asp Arg Leu
110 115 120

Thr Arg Arg Pro Pro Arg Asp Ala His Tyr Arg Met Leu Asn Glu
125 130 135

Tyr Gly Pro Thr Glu Asn Ala Val Asp Thr Thr Trp Ala Ile Val
140 145 150

Ser Pro Ala Ser Glu His Gly Ser Glu Leu Pro Ser Ile Gly Arg
155 160 165

Pro Leu Asp Asn Val Gln Val Tyr Val Leu Asp Ser Ala Leu Glu
170 175 180

Pro Val Pro Ala Arg Val Ile Gly Glu Leu Tyr Ile Gly Gly Glu
185 190 195

Asn Leu Ala Arg Gly Tyr Val Asp Asp Ala Ala Leu Thr Gly Ala
200 205 210

Arg Phe Val Pro Asp Pro Phe Gly Ala Pro Gly Ser Arg Leu Tyr
215 220 225

Arg Thr Gly Asp Leu Val Arg Trp Leu Pro Asp Gly Asn Leu Asp
230 235 240

Phe Ile Gly Arg Ala Asp His Gln Val Lys Ile Arg Gly Phe Arg
245 250 255

Ile Glu Leu Gly Glu Ile Glu Ala Ala Leu Ala Asp His Pro Gly

260 265 270

Val Glu Thr Ala Val Val Leu Ala Arg Gln Glu Arg Ala Gly Asp
275 280 285

Lys Arg Leu Val Ala Tyr Trp Ser Pro Arg Leu Ala Ser Val Ala
290 295 300

Ala Ser Glu Leu Arg Asp His Leu Arg Thr Arg Leu Pro Glu Tyr
305 310 315

Met Ile Pro Ala Ala Phe Val Glu Leu Arg Glu Met Pro Arg Met
320 325 330

Ala Ser Gly Lys Ile Asp Leu Lys Val Leu Pro Ala Pro Pro Asp
335 340 345

Asp Ser Ser Asp Arg Thr Asp Tyr Asp Pro Pro Ser Thr Pro Val
350 355 360

Glu Val Lys Leu Ala Arg Ile Trp Ser Gly Leu Leu Gly Val Glu
365 370 375

Arg Val Ser Arg Thr Asp Ser Phe Phe Ala Pro Gly Gly Pro Ser
380 385 390

(2) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 472
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCGGCGGGT TCCAGACGGC CATGGTGCTG ACGACGGGAC GGGACAATGA 50
GAAGTAGCGT CGCGGTCACC GGCATCGGCC TGGTGGCCGC CAACGGGCTC 100
ACCACCGAGG ACGTGTGGTC GGCCGTGCTC GGCGGCCGCA GCGGCCTTGG 150
AACGATCACC CGTTTCGACG CCGCGGGCTA CCCGGCCCGG ATCGCCGGCG 200
AGGTGTCGCA GTTCGTGGCC GAGGAGCACA TCGCCGACCG GCTGATCCCG 250
CAGACCGACC ACATGACCCG GCTGGCGCTG GCCGCGGCCG AGTCGGCGAT 300
CCGGGACGCC AAGGTGGGAC CTGGCCGAGC TGCCCGATTC GGCGCGGGCG 350
TGGTCACCGC CGCGACGGCA GGCGGCTTCG AGTTCGGCCA GCGGGAGCTG 400
GAGAACCTGT GGCGCAAGGG GCCTGAGCAC GTCAGCCCCT ACCAGTCCTT 450
CGCCTGGTTC TACGCCGTCA AC 472

(2) INFORMATION FOR SEQ ID NO:20:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 142
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Arg Ser Ser Val Ala Val Thr Gly Ile Gly Leu Val Ala Ala
5 10 15

Asn Gly Leu Thr Thr Glu Asp Val Trp Ser Ala Val Leu Gly Gly
20 25 30

Arg Ser Gly Leu Gly Thr Ile Thr Arg Phe Asp Ala Ala Gly Tyr
35 40 45

Pro Ala Arg Ile Ala Gly Glu Val Ser Gln Phe Val Ala Glu Glu
50 55 60

His Ile Ala Asp Arg Leu Ile Pro Gln Thr Asp His Met Thr Arg
65 70 75

Leu Ala Leu Ala Ala Ala Glu Ser Ala Ile Arg Asp Ala Lys Val
80 85 90

Gly Pro Gly Arg Ala Ala Arg Phe Gly Ala Gly Val Val Thr Ala
95 100 105

Ala Thr Ala Gly Gly Phe Glu Phe Gly Gln Arg Glu Leu Glu Asn
110 115 120

Leu Trp Arg Lys Gly Pro Glu His Val Ser Pro Tyr Gln Ser Phe
125 130 135

Ala Trp Phe Tyr Ala Val Asn
140

(2) INFORMATION FOR SEQ ID NO:21:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 637
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TATATTACTC CAGGTTGCTT ACGAAGCATT GGAGATGTCC GGATATTTCG 50
CCGATTCGTC CAGGCCTGAG GATGTCGGTT GCTATATTGG AGCTTGTGCA 100
ACAGATTACG ATTTCAACGT AGCATCCCAT CCTCCCACGG CGTATTCAGC 150
GACTGGCACG CTCCGATCTT TTCTAAGTGG CAAGCTGTCG CATTACTTTG 200
GTTGGTCCGG TCCCTCTCTT GTCCTAGACA CTGCCTGCTC TTCGTCGGCG 250
GTGGCTATTC ATACTGCATG TACTGCTTTG AGGACTGGCC AGTGTTCTCA 300
AGCTCTAGCA GGCGGGATCA CGTTGATGAC AAGCCCGTAT CTCTATGAGA 350
ACTTCTCTGC AGCCCATTTC TTGAGTCCAA CGGGAGGTTC AAAGCCGTTC 400

AGCGCAGRTG CAGATGGATA CTGTAGAGGA GAAGGTGGTG GCCTCGTGGT 450
CTTGAAACGA CTTTCAGATG CTCTCAGGGA TGATGACCAT ATTATTAGTG 500
TCATCGCTGG CTCGGCGGTC AACCAGAACG ACAACTGCGT GCCTATCACC 550
GTCCCTCACA CTTCGTCTCA GGGAAATCTC TATGAACGAG TTACCAGACA 600
GGCAGGGGTG ACACCCAATA AAGTCACTTT TGTGGAA 637

(2) INFORMATION FOR SEQ ID NO:22:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ile Leu Leu Gln Val Ala Tyr Glu Ala Leu Glu Met Ser Gly Tyr
5 10 15

Phe Ala Asp Ser Ser Arg Pro Glu Asp Val Gly Cys Tyr Ile Gly
20 25 30

Ala Cys Ala Thr Asp Tyr Asp Phe Asn Val Ala Ser His Pro Pro
35 40 45

Thr Ala Tyr Ser Ala Thr Gly Thr Leu Arg Ser Phe Leu Ser Gly
50 55 60

Lys Leu Ser His Tyr Phe Gly Trp Ser Gly Pro Ser Leu Val Leu
65 70 75

Asp Thr Ala Cys Ser Ser Ser Ala Val Ala Ile His Thr Ala Cys
80 85 90

Thr Ala Leu Arg Thr Gly Gln Cys Ser Gln Ala Leu Ala Gly Gly
95 100 105

Ile Thr Leu Met Thr Ser Pro Tyr Leu Tyr Glu Asn Phe Ser Ala
110 115 120

Ala His Phe Leu Ser Pro Thr Gly Gly Ser Lys Pro Phe Ser Ala
125 130 135

Xaa Ala Asp Gly Tyr Cys Arg Gly Glu Gly Gly Gly Leu Val Val
140 145 150

Leu Lys Arg Leu Ser Asp Ala Leu Arg Asp Asp Asp His Ile Ile
155 160 165

Ser Val Ile Ala Gly Ser Ala Val Asn Gln Asn Asp Asn Cys Val
170 175 180

Pro Ile Thr Val Pro His Thr Ser Ser Gln Gly Asn Leu Tyr Glu
185 190 195

Arg Val Thr Arg Gln Ala Gly Val Thr Pro Asn Lys Val Thr Phe
200 205 210

Val Glu

(2) INFORMATION FOR SEQ ID NO:23:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1177
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCACGACGGG CAAGCCCAAG GGGGGCGATG AACAGCCATC GAGGAATTTG 50
CAATCGCTTA CTGTGGATGC AAGATGCTTA CAAACTAACT GAAACTGATC 100
GCGTTCTGCA AAAAACGCCT TTTAGTTTCG ACGTTTCCGT TTGGGAGTTT 150
TTCTGGCCTC TCTTGACAGG GGCGCGTTTA GTGATGGCTC AACCAGGCGG 200
ACAGCGAGAT GCAACTTACT TAATTAACAC CATCGTCCAA GAGGAAATTA 250
CAACACTGCA TTTTGTCCCC TCCATGTTGC GGATATTTCT CCAAACTAAA 300
GGGCTAGAAC GTTGTCAATC TCTAAAACGG GTGTTTTGTA GTGGAGAAGC 350
CTTACCAGTT GACCTCCAGG AGCGGTTTTT TGACTCGATG GGATGTGAAC 400
TACACAACCT CTATGGTCCT ACCGAAGCGG CAATTGATGT CACATTTTGG 450
CAGTGTCAAA GAGAGAGTAA CTTAAAAAGT GTACCGATTG GGAGAGCGAT 500
CGCCAACACT CAAMTTTATA TCCTCGACTC CCATTTACAA GCAGTTCCCT 550
TGGGTGCGAT CGGCGAACTT TATATTGGTG GTATCGGCGT TGCTAGAGGS 600
TATCTTAACC GTCCAGACTT AACAGCCGAG CGATTTATTT CCCATCCCTT 650
TAAGGAAGGC GRRAAACTTT ACAAAACAGG AGACTTAGCC CGATATCTGG 700
CCGATGGCAA TATCGAATAC ATCGGTAGAA TTGATCATCA AGTAAAAATT 750
CGGGGTTTCC GCATCGAACT TGGAGAAATC GAAACTTTAC TAGCACAACA 800
CCCGACCATA CAGCAAACTG TCGTCACAGC TAGAATTGAT CATCTCGAAA 850
ACCAGCGATT AGTCGCCTAC ATCGTTCCTC ATTCAGAGCA GACACTAACC 900
ACAGACGAAC TGCGCCACTT CCTCAAAAAG AAACTGCCAG AATATATGGT 950
GCCTAGTACT TTCGTTTTCC TAGACACTCT ACCCCTAACC CCCAACGGCA 1000
AAATTGACCG TCGCGCTTTA CCAGCACCCG ACTCAACAAG GCTTGATTCA 1050
GAAAACACAT ATCTTGCTCC CCGCGATTAA TTAGAATTTC AGTTGACTAA 1100
AATTTGGTCA GAAATTTTAG GTATCCAGCC TATCGGTGTC AGGGACAACT 1150
TCTTCTTCCT TGGGCGGCCC CTCCCTT 1177

(2) INFORMATION FOR SEQ ID NO:24:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Ala Arg Arg Ala Ser Pro Arg Gly Ala Met Asn Ser His Arg Gly
5 10 15

Ile Cys Asn Arg Leu Leu Trp Met Gln Asp Ala Tyr Lys Leu Thr
20 25 30

Glu Thr Asp Arg Val Leu Gln Lys Thr Pro Phe Ser Phe Asp Val
35 40 45

Ser Val Trp Glu Phe Phe Trp Pro Leu Leu Thr Gly Ala Arg Leu
50 55 60

Val Met Ala Gln Pro Gly Gly Gln Arg Asp Ala Thr Tyr Leu Ile
65 70 75

Asn Thr Ile Val Gln Glu Glu Ile Thr Thr Leu His Phe Val Pro
80 85 90

Ser Met Leu Arg Ile Phe Leu Gln Thr Lys Gly Leu Glu Arg Cys
95 100 105

Gln Ser Leu Lys Arg Val Phe Cys Ser Gly Glu Ala Leu Pro Val
110 115 120

Asp Leu Gln Glu Arg Phe Phe Asp Ser Met Gly Cys Glu Leu His
125 130 135

Asn Leu Tyr Gly Pro Thr Glu Ala Ala Ile Asp Val Thr Phe Trp
140 145 150

Gln Cys Gln Arg Glu Ser Asn Leu Lys Ser Val Pro Ile Gly Arg
155 160 165

Ala Ile Ala Asn Thr Gln Xaa Tyr Ile Leu Asp Ser His Leu Gln
170 175 180

Ala Val Pro Leu Gly Ala Ile Gly Glu Leu Tyr Ile Gly Gly Ile
185 190 195

Gly Val Ala Arg Gly Tyr Leu Asn Arg Pro Asp Leu Thr Ala Glu
200 205 210

Arg Phe Ile Ser His Pro Phe Lys Glu Gly Gly Lys Leu Tyr Lys
215 220 225

Thr Gly Asp Leu Ala Arg Tyr Leu Ala Asp Gly Asn Ile Glu Tyr
230 235 240

Ile Gly Arg Ile Asp His Gln Val Lys Ile Arg Gly Phe Arg Ile
245 250 255

Glu Leu Gly Glu Ile Glu Thr Leu Leu Ala Gln His Pro Thr Ile
260 265 270

Gln Gln Thr Val Val Thr Ala Arg Ile Asp His Leu Glu Asn Gln
275 280 285

Arg Leu Val Ala Tyr Ile Val Pro His Ser Glu Gln Thr Leu Thr

290 295 300

Thr Asp Glu Leu Arg His Phe Leu Lys Lys Lys Leu Pro Glu Tyr
305 310 315

Met Val Pro Ser Thr Phe Val Phe Leu Asp Thr Leu Pro Leu Thr
320 325 330

Pro Asn Gly Lys Ile Asp Arg Arg Ala Leu Pro Ala Pro Asp Ser
335 340 345

Thr Arg Leu Asp Ser Glu Asn Thr Tyr Leu Ala Pro Arg Asp Xaa
350 355 360

Leu Glu Phe Gln Leu Thr Lys Ile Trp Ser Glu Ile Leu Gly Ile
365 370 375

Gln Pro Ile Gly Val Arg Asp Asn Phe Phe Leu Gly Arg Pro
380 385 390

Leu Pro

(2) INFORMATION FOR SEQ ID NO:25:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 406
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met Ser Ile Arg Thr Val Val Thr Gly Leu Gly Ile Ala Ala Pro
5 10 15

Asn Gly Leu Gly Ile Glu Glu Tyr Trp Ser Ala Thr Leu Ala Gly
20 25 30

Arg Gly Ala Ile Gly Pro Leu Thr Arg Phe Asp Ala Ser Ser Tyr
35 40 45

Pro Ser Arg Leu Ala Gly Glu Ile Arg Gly Phe Thr Ala Ala Glu
50 55 60

His Leu Pro Gly Arg Leu Leu Pro Gln Thr Asp Arg Met Thr Gln
65 70 75

Leu Ala Leu Val Ser Ala Gly Trp Ala Leu Asp Asp Ala Gly Val
80 85 90

Val Pro Asp Glu Leu Pro Ala Tyr Asp Met Gly Val Ile Thr Ala
95 100 105

Ser His Ala Gly Gly Phe Glu Phe Gly Gln Asn Glu Leu Lys Ala
110 115 120

Leu Trp Ser Lys Gly Gly Lys Tyr Val Ser Ala Tyr Gln Ser Phe
125 130 135

Ala Trp Phe Tyr Ala Val Asn Ser Gly Gln Ile Ser Ile Arg Asn
140 145 150

Gly Met Arg Gly Pro Ser Gly Val Val Val Ser Asp Gln Ala Gly
155 160 165

Gly Leu Asp Ala Leu Ala Gln Ala Arg Arg Gln Ile Arg Lys Gly
170 175 180

Thr Pro Leu Ile Val Ser Gly Ala Val Asp Ala Ser Leu Cys Thr
185 190 195

Trp Gly Trp Val Ala Gln Leu Ala Gly Gly Arg Leu Ser Arg Ser
200 205 210

Asp Asp Pro Gly His Ala Tyr Val Pro Phe Asp Asp Ala Ala Val
215 220 225

Gly His Val Pro Gly Glu Gly Gly Ala Leu Leu Ile Leu Glu Glu
230 235 240

Ala Glu His Ala Arg Ser Arg Gly Ala Arg Arg Ile Tyr Gly Glu
245 250 255

Ile Thr Gly His Ala Ser Thr Phe Asp Pro Pro Pro Trp Ser Gly
260 265 270

Arg Gly Pro Ala Val Gln Arg Val Ile Glu Glu Ala Leu Ala Asp
275 280 285

Ala Gly Thr Val Pro Asp Glu Val Asp Val Val Phe Ala Asp Ala
290 295 300

Ala Ala Leu Pro Glu Leu Asp Arg Ile Glu Ala Ala Ala Ile Thr
305 310 315

Lys Val Phe Gly Pro His Ala Val Pro Val Thr Ala Pro Lys Thr
320 325 330

Met Thr Gly Arg Leu Tyr Ser Gly Ala Ala Pro Leu Asp Val Ala
335 340 345

Ala Ala Cys Leu Ala Ile Arg Asp Gly Leu Ile Pro Pro Thr Ile
350 355 360

His Ser Ser Leu Ser Gly Arg Tyr Glu Ile Asp Leu Val Thr Gly
365 370 375

Ala Pro Arg Thr Ala Pro Val Arg Thr Ala Leu Val Val Ala Arg

380 385 390

Gly His Gly Gly Phe Asn Ser Ala Val Val Val Arg Ala Pro Arg
395 400 405

Asp

(2) INFORMATION FOR SEQ ID NO:26:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Thr Ser Glu Leu Leu Glu Arg Thr Ala Val Arg Ser Ala Thr
5 10 15

Ala Val Phe Thr Gly Ile Gly Val Thr Ala Pro Asn Gly Leu Gly
20 25 30

Thr Ala Ala Trp Trp Gln Ala Thr Val Ala Gly Glu Ser Gly Ile
35 40 45

Arg Pro Val Ser Arg Phe Asp Ala Ser Gly Tyr Pro Ser Thr Leu
50 55 60

Ala Gly Glu Val Pro Gly Phe Asp Ala Glu Glu His Ile Pro Ser
65 70 75

Arg Leu Leu Ser Gln Thr Asp His Met Thr Arg Leu Ala Leu Thr
80 85 90

Ala Ala Lys Glu Ala Leu Glu Asp Ser Gly Ala Asp Pro Ala Glu
95 100 105

Met Pro Gln Tyr Ser Ala Gly Ala Val Thr Ala Ala Ser Ala Gly
110 115 120

Gly Phe Glu Phe Gly Gln Arg Glu Leu Gln Ala Leu Trp Ser Lys
125 130 135

Gly Gly Gln Tyr Val Ser Ala Tyr Gln Ser Tyr Ala Trp Phe Tyr
140 145 150

Ala Val Asn Thr Gly Gln Ile Ser Ile Arg His Gly Leu Arg Gly
155 160 165

Pro Ser Gly Val Leu Val Thr Glu Gln Ala Gly Gly Leu Glu Ala
170 175 180

```
Val Ala Gln Ala Arg Arg Gln Leu Arg Lys Gly Ser Lys Leu Ile
                185                 190                 195

Val Thr Gly Gly Val Asp Gly Ala Val Cys Pro Trp Gly Trp Thr
            200                 205                 210

Ala Gln Leu Ala Gly Gly Arg Met Ser Pro Val Ala Asp Pro Ala
        215                 220                 225

Arg Ala Phe Leu Pro Phe Asp Ser Glu Ala Ser Gly Tyr Val Ala
    230                 235                 240

Gly Glu Gly Gly Ala Ile Leu Val Leu Glu Asp Ala Glu Ala Ala
245                 250                 255

Arg Glu Arg Gly Ala Arg Ile Tyr Gly Arg Leu Ser Gly Tyr Ala
            260                 265                 270

Ala Thr Phe Asp Pro Ala Pro Gly Arg Gly Gly Glu Pro Gly Leu
        275                 280                 285

Arg Arg Ala Ala Glu Leu Ala Leu Thr Glu Ala Gly Leu Ser Ala
    290                 295                 300

Ser Asp Val Asp Val Val Phe Ala Asp Ala Ser Gly Val Pro Glu
305                 310                 315

Leu Asp Arg Gln Glu Glu Ala Ala Leu Thr Ala Leu Phe Gly Pro
            320                 325                 330

Arg Gly Val Pro Val Thr Ala Pro Lys Thr Met Thr Gly Arg Leu
        335                 340                 345

Ser Ala Gly Gly Ala Ser Leu Asp Leu Ala Ala Ala Leu Leu Ser
    350                 355                 360

Ile Arg Asp Ala Val Ile Pro Pro Thr Val Asn Val Thr Ser Pro
365                 370                 375

Val Ala Ala Asp Ala Leu Asp Leu Val Thr Glu Ala Arg Arg Gly
            380                 385                 390

Pro Val Arg Thr Ala Leu Val Leu Ala Arg Gly Thr Gly Gly Phe
        395                 400                 405

Asn Ala Ala Ala Val Val Thr Ala Ala Asn
    410                 415
```

(2) INFORMATION FOR SEQ ID NO:27:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein

(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met Ile Pro Val Ala Val Thr Gly Met Gly Val Ala Ala Pro Asn
5 10 15

Gly Leu Gly Ala Ala Asp Tyr Trp Ala Ala Thr Arg Gly Gly Lys
20 25 30

Ser Gly Ile Gly Arg Ile Thr Arg Phe Asp Pro Ser Ser Tyr Pro
35 40 45

Ala Arg Leu Ala Gly Glu Ile Pro Gly Phe Glu Ala Ala Glu His
50 55 60

Leu Pro Gly Arg Leu Leu Pro Gln Thr Asp Arg Val Thr Arg Leu
65 70 75

Ser Leu Ala Ala Ala Asp Trp Ala Leu Ala Asp Ala Gly Val Glu
80 85 90

Pro Glu Ser Phe Asp Pro Leu Asp Met Gly Val Val Thr Ala Gly
95 100 105

His Ala Gly Gly Phe Glu Phe Gly Gln Gly Glu Leu Gln Lys Leu
110 115 120

Trp Ala Lys Gly Ser Gln Phe Val Ser Ala Tyr Gln Ser Phe Ala
125 130 135

Trp Phe Tyr Ala Val Asn Ser Gly Gln Ile Ser Ile Arg His Gly
140 145 150

Met Lys Gly Pro Asn Gly Val Val Val Ser Asp Gln Ala Gly Gly
155 160 165

Leu Asp Ala Leu Ala Gln Ala Arg Arg Leu Val Arg Lys Gly Thr
170 175 180

Pro Leu Ile Val Cys Gly Ala Val Asp Ala Ser Ile Cys Pro Trp
185 190 195

Gly Trp Val Ala Gln Leu Ala Gly Gly Arg Met Ser Asp Ser Asp
200 205 210

Glu Pro Ala Arg Ala Tyr Leu Pro Phe Asp Arg Asp Ala Arg Gly
215 220 225

Tyr Leu Pro Gly Glu Gly Gly Ala Ile Leu Ile Met Glu Pro Ala
230 235 240

Ala Ala Ala Arg Ala Arg Gly Ala Lys Val Tyr Gly Glu Ile Ser
245 250 255

Gly Tyr Gly Ala Thr Phe Asp Pro Pro Pro Gly Ser Gly Ser Gly

```
                260                 265                 270

Ser Thr Leu Arg Thr Ala Ile Arg Val Ala Leu Asp Asp Ala Gly
                275                 280                 285

Val Ala Pro Gly Asp Val Asp Ala Val Phe Ala Asp Gly Ala Gly
                290                 295                 300

Val Pro Glu Leu Asp Arg Ala Glu Ala Glu Ala Ile Thr Asp Val
                305                 310                 315

Phe Gly Ser Gly Gly Val Pro Val Thr Val Pro Lys Thr Met Thr
                320                 325                 330

Gly Arg Leu Tyr Ser Gly Ala Ala Pro Leu Asp Val Ala Cys Ala
                335                 340                 345

Leu Leu Ala Met Gln Ala Gly Val Ile Pro Pro Thr Val His Ile
                350                 355                 360

Asp Pro Cys Pro Glu Tyr Gly Leu Asp Leu Val Leu His Gln Ala
                365                 370                 375

Arg Pro Ala Thr Val Arg Thr Ala Leu Val Leu Ala Arg Gly His
                380                 385                 390

Gly Gly Phe Asn Ser Ala Met Ala Val Arg Ala Gly Arg
                395                 400
```

(2) INFORMATION FOR SEQ ID NO:28
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 407
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28

```
Met Ser Ala Arg Phe Leu Val Thr Gly Ile Gly Val Ala Ala Pro
                 5                  10                  15

Ser Gly Leu Gly Val Glu Asp Phe Trp Ser Val Thr Arg Ile Gly
                 20                  25                  30

Lys Asn Ala Ile Gly Pro Val Thr Arg Phe Asp Ala Ser Ala Tyr
                 35                  40                  45

Pro Ser Arg Leu Ala Gly Glu Ile His Gly Phe Glu Pro Lys Glu
                 50                  55                  60

His Leu Pro Gly Arg Leu Val Pro Gln Thr Asp Arg Val Thr Gln
                 65                  70                  75
```

Leu Ala Leu Val Ala Ala Asp Cys Ala Phe Ala Asp Ala Gly Ile
80 85 90

Glu Pro Gly Thr Ile Asp Pro Tyr Ala Met Gly Val Val Thr Ala
95 100 105

Ala Gly Ala Gly Gly Phe Glu Phe Ala Glu Asn Glu Leu Arg Lys
110 115 120

Leu Trp Ser Glu Gly Ala Lys Arg Val Ser Ala Tyr Gln Ser Phe
125 130 135

Ala Trp Phe Tyr Ala Val Asn Ser Gly Gln Ile Ser Ile Arg Asn
140 145 150

Gly Leu Arg Gly Pro Ala Gly Val Val Ile Ser Asp Gln Ala Gly
155 160 165

Gly Leu Asp Ala Leu Ala Gln Ala Arg Arg Gln Leu Arg Lys Gly
170 175 180

Ser Lys Leu Ile Ala Thr Gly Gly Phe Asp Ala Pro Ile Cys Ser
185 190 195

Leu Gly Trp Ala Ser Gln Pro Arg Thr Gly Gly Leu Met Phe His
200 205 210

Glu Arg Thr Glu Pro Glu Arg Ala Tyr Leu Pro Phe Glu Asp Ala
215 220 225

Ala Ala Gly Tyr Val Pro Gly Glu Gly Gly Ala Met Leu Ile Leu
230 235 240

Glu Asp Glu Asp Ser Ala Arg Asp Arg Gly Ala Arg Thr Val Tyr
245 250 255

Gly Glu Phe Ala Gly Tyr Gly Ala Thr Leu Asp Pro Lys Pro Gly
260 265 270

Ser Gly Arg Glu Pro Gly Leu Arg Arg Ala Ile Asp Val Ala Leu
275 280 285

Thr Asp Ala Ala Cys His Pro Ala Glu Val Glu Val Val Phe Ala
290 295 300

Asp Gly Ala Ala Thr Pro Arg Leu Asp Arg Glu Glu Ala Glu Ala
305 310 315

Ile Thr Ala Val Phe Gly Pro Arg Ala Val Pro Val Thr Val Pro
320 325 330

Lys Thr Met Thr Gly Arg Ile Asn Ser Gly Gly Ala Pro Ile Asp
335 340 345

Val Val Ser Ala Val Leu Ser Met Arg Glu Gly Leu Ile Pro Pro

350 355 360

Thr Thr Asn Val Glu Leu Ser Asp Ala Tyr Asp Leu Asp Leu Val
365 370 375

Ala Val Arg Pro Arg Thr Ala Ser Val Arg Thr Ala Leu Val Leu
380 385 390

Ala Arg Gly Arg Gly Gly Phe Asn Ser Ala Val Val Val Arg Ala
395 400 405

Val Asp

(2) INFORMATION FOR SEQ ID NO:29:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:643
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGATCTGCTT GAGGTAGTCT ACGAGGCACT GGAGTCAGCA GGGTACTTTG 50
GCGCCAAGTC AAACCCGGAA CCTGATGACT ATGGATGCTA TATCGGTGCA 100
GTGATGAACA ACTACTATGA CAACGTTTCT TGCCATCCAC CCACCGCATA 150
CGCTACTCTT GGAACGTCGC GTTGCTTCCT TAGTGGCTGC ATGAGCCATT 200
ACTTTGGATG GACGGGACCT TCCTTGACCA TTGATACGGC TTGCTCGTCA 250
TCACTAGTTG CTATAAACAC CGCTTGTAGA GCAATATGGT CTGGTGAGTG 300
CTCCCGGGCC ATAGCTGGGG GTACCAATGT CTTCACAAGT CCGTTTGACT 350
ACCAGAATCT TCGCGCCGCA GGATTCCTCA GCCCTAGCGG GCAATGCAAG 400
CCGTTTGATG CTTCTGCTGA TGGCTACTGC CGTGGAGAAG GAGTTGGTGT 450
CGTTGTGCTT AAGCCTTTGA CGGCTGCTAT GCAAGAGAAC GATAACATCC 500
TTGGCGTCAT TGTGGGGTCT GCAGCAAACC AAAACCAAAA CCTCAGTCAT 550
ATCACGGTGC CCCATTCGGG CTCACAAGTC CAGCTTTATC GAAAGGTGAT 600
GAAGCTTGCA GGTATAGAGC CAGAGTCAGT CTCCTACGTT GAG 643

(2) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ile Leu Leu Gln Val Ala Tyr Glu Ala Leu Glu Met Ser Gly Tyr
5 10 15

Phe Ala Asp Ser Ser Arg Pro Glu Asp Val Gly Cys Tyr Ile Gly
20 25 30

Ala Cys Ala Thr Asp Tyr Asp Phe Asn Val Ala Ser His Pro Pro

35 40 45

Thr Ala Tyr Ser Ala Thr Gly Thr Leu Arg Ser Phe Leu Ser Gly
50 55 60

Lys Leu Ser His Tyr Phe Gly Trp Ser Gly Pro Ser Leu Val Leu
65 70 75

Asp Thr Ala Cys Ser Ser Ser Ala Val Ala Ile His Thr Ala Cys
80 85 90

Thr Ala Leu Arg Thr Gly Gln Cys Ser Gln Ala Leu Ala Gly Gly
95 100 105

Ile Thr Leu Met Thr Ser Pro Tyr Leu Tyr Glu Asn Phe Ser Ala
110 115 120

Ala His Phe Leu Ser Pro Thr Gly Gly Ser Lys Pro Phe Ser Ala
125 130 135

Xaa Ala Asp Gly Tyr Cys Arg Gly Glu Gly Gly Gly Leu Val Val
140 145 150

Leu Lys Arg Leu Ser Asp Ala Leu Arg Asp Asp Asp His Ile Ile
155 160 165

Ser Val Ile Ala Gly Ser Ala Val Asn Gln Asn Asp Asn Cys Val
170 175 180

Pro Ile Thr Val Pro His Thr Ser Ser Gln Gly Asn Leu Tyr Glu
185 190 195

Arg Val Thr Arg Gln Ala Gly Val Thr Pro Asn Lys Val Thr Phe
200 205 210

Val Glu

(2) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:643
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AATCCTCATG GAATCAGCTT GGCAAACACT AGAAAACGCT GGCATAACTG 50
CGAACAAAGT AGCTGGCAGC AGTACAGGAG TTTTTGTGGG TGCTAGTGGC 100
TCTGATTACT GTTGGGTAAT GGAGCGGGTA GGTATTCCCA TAGAAGCTCA 150
CGTTGCAACG GGCACGTCGT TGGCAGCGCT GGCAAATCGC ATCTCTTACT 200
TTTTTGACTT GCGAGGCCCA AGCATCGTCA TTGATACGGC GTGTTCTAGT 250
TCGTTGATGG CAGTGCATCA GGCGGTTCAA TCTATCCGAG CAGGTGAGTG 300

```
CTTACAAGCA CTGGTGGGCG GTATACATAT CATGAGCCAT CCGGCTAACA  350
GTATTGCATA TTACAAGGCT GGGATGTTGG CGCATGATGG CAAGTGCAAG  400
ACATTTGACG ATCGCGCAGA TGGGTACGTT CGCAGTGAAG GCGCTGTGAT  450
GCTTCTGCTC AAGCAATTGC ATCAGGCGGA AGCAGATGGC GATCTAATTT  500
ATGCGACAAT CAAGGGGTCA GCCTCGAATC ATGGTGGACA GTCCGCCGGC  550
CTCACCGTAC CGAATCCGCA ACAGCAGGCA GCACTCTTAA CCAATGCCTG  600
GAAAGCCTCT GGTGTAGACC CTAACACGAT TAGTTTTATC GAA         643
```

(2) INFORMATION FOR SEQ ID NO:32:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Ile Leu Met Glu Ser Ala Trp Gln Thr Leu Glu Asn Ala Gly Ile
                  5                  10                  15

Thr Ala Asn Lys Val Ala Gly Ser Ser Thr Gly Val Phe Val Gly
             20                  25                  30

Ala Ser Gly Ser Asp Tyr Cys Trp Val Met Glu Arg Val Gly Ile
         35                  40                  45

Pro Ile Glu Ala His Val Ala Thr Gly Thr Ser Leu Ala Ala Leu
     40                  55                  60

Ala Asn Arg Ile Ser Tyr Phe Phe Asp Leu Arg Gly Pro Ser Ile
 65                  70                  75

Val Ile Asp Thr Ala Cys Ser Ser Ser Leu Met Ala Val His Gln
             80                  85                  90

Ala Val Gln Ser Ile Arg Ala Gly Glu Cys Leu Gln Ala Leu Val
         95                 100                 105

Gly Gly Ile His Ile Met Ser His Pro Ala Asn Ser Ile Ala Tyr
    110                 115                 120

Tyr Lys Ala Gly Met Leu Ala His Asp Gly Lys Cys Lys Thr Phe
125                 130                 135

Asp Asp Arg Ala Asp Gly Tyr Val Arg Ser Glu Gly Ala Val Met
            140                 145                 150

Leu Leu Leu Lys Gln Leu His Gln Ala Glu Ala Asp Gly Asp Leu
        155                 160                 165

Ile Tyr Ala Thr Ile Lys Gly Ser Ala Ser Asn His Gly Gly Gln
    170                 175                 180
```

Ser Ala Gly Leu Thr Val Pro Asn Pro Gln Gln Gln Ala Ala Leu
185 190 195

Leu Thr Asn Ala Trp Lys Ala Ser Gly Val Asp Pro Asn Thr Ile
200 205 210

Ser Phe Ile Glu

(2) INFORMATION FOR SEQ ID NO:33:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:637
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TATATTACTC CAGGTTGCTT ACGAAGCATT GGAAATGTCC GGGTATTTCG 50
CCGACTCGTC CAAGCCTGAG GACGTAGGTT GCTATATTGG AGCTTGTGCA 100
ACAGATTACG ATTTCAGCGT AGCGTCCCAT CCTCCTACGG CATACTCAGC 150
AACTGGCACG CTCCGATCTT TCCTGAGTGG CAAGCTGTCA CATTACTTTG 200
GTTGGTCTGG TCCCTCTCTT GTCCTGGACA CCGCCTGCTC TTCATCGGCG 250
GTGGCCATTC ACACTGCATG TACTGCTTTG AGGACTGGCC AGTGTTCTCA 300
GGCTTTAGCA GGCGGGATTA CTTTGATGAC CAGCCCGTAT CTCTTTGAGA 350
ACTTTGCTGC CGCCCATTTC TTGAGCCCAA CGGGAGGCTC AAAGCCGTTC 400
AGTGCAGATG CAGATGGGTA TTGTAGAGGA GAAGGGGGTG GGCTCGTGGT 450
CTTGAAACGA CTTTCAGATG CTATCAGGGA TAACGACCAC ATCATTAGCG 500
TCATCGCTGG CTCAGCCGTC AACCAGAACG CTAACTGTGT GCCTATCACC 550
GTCCCTCATA CTTCGTCTCA GGGCAATCTC TATGAACGAG TTACCGCACA 600
GGCAGGGGTG ACACCTAATA AGGTCACTTT TGTGGAA 637

(2) INFORMATION FOR SEQ ID NO:34:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ile Leu Leu Gln Val Ala Tyr Glu Ala Leu Glu Met Ser Gly Tyr
5 10 15

Phe Ala Asp Ser Ser Lys Pro Glu Asp Val Gly Cys Tyr Ile Gly
20 25 30

Ala Cys Ala Thr Asp Tyr Asp Phe Ser Val Ala Ser His Pro Pro
35 40 45

Thr Ala Tyr Ser Ala Thr Gly Thr Leu Arg Ser Phe Leu Ser Gly
50 55 60

Lys Leu Ser His Tyr Phe Gly Trp Ser Gly Pro Ser Leu Val Leu
65 70 75

Asp Thr Ala Cys Ser Ser Ser Ala Val Ala Ile His Thr Ala Cys
80 85 90

Thr Ala Leu Arg Thr Gly Gln Cys Ser Gln Ala Leu Ala Gly Gly
95 100 105

Ile Thr Leu Met Thr Ser Pro Tyr Leu Phe Glu Asn Phe Ala Ala
110 115 120

Ala His Phe Leu Ser Pro Thr Gly Gly Ser Lys Pro Phe Ser Ala
125 130 135

Asp Ala Asp Gly Tyr Cys Arg Gly Glu Gly Gly Gly Leu Val Val
140 145 150

Leu Lys Arg Leu Ser Asp Ala Ile Arg Asp Asn Asp His Ile Ile
155 160 165

Ser Val Ile Ala Gly Ser Ala Val Asn Gln Asn Ala Asn Cys Val
170 175 180

Pro Ile Thr Val Pro His Thr Ser Ser Gln Gly Asn Leu Tyr Glu
185 190 195

Arg Val Thr Ala Gln Ala Gly Val Thr Pro Asn Lys Val Thr Phe
200 205 210

Val Glu

(2) INFORMATION FOR SEQ ID NO:35:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:691
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCATCTGCTA GAAATCAGCT ACGAGGCGCT CGAGAATGCA GGCTTTCCAC 50
TGCCTAGCAT TGCTGGCACG AACATGGGTG TCTTTGTCGG CGGAAGCAAC 100
TCTGAGTATC GAGCGCACAT CGGAAACGAT ACCGACAACT TACCGATGTT 150
TGAAGCAACA GGCAATGCAG AATCTCTGCT GGCGAATCGA GTCTCTTATG 200
TGTATGATCT CCACGGCGCA AGTCTGACGA TTGGTACCGC TTGTTCCGTC 250
GAGTTTAGCA GCTTTGGATA GCGCGTTTCT CAGCTTGCAG CTGGTAAGTC 300
GTCCACAGCA ATTGTTGCCG GCTCCGTTGT TCGAATCGTA CCGTCATCGA 350
CCATCTCACC TTCTACTATG AAGTAAGCAG TCATGGCTCT TGACACGGAG 400
ACTACTCACC ATTCCAGGCT TCTGTCACCA GAAGGGCGGT GTTATGCGTT 450
CGATGACAGA GCCACTAGTG GTTTTGGAAG GGGTGAAGGT TCTGCCTGCA 500
TAATATTGGA AACCTTAGAG GCAGCCTTAA GAGACAACGA CCCAATCCGA 550
TCGGTCATTC GCAATTCGGG AGTCAATCAA GATGGTAAAA CTGCAGGTAT 600

CACAATGCCA AATGGGGAAG CGCAAGCTTC ATTGATACAA TCTGTTTATC 650
GCACTGCTGG ATTGGACCCT CTGCAGACAG ATTACGTCGA G 691

(2) INFORMATION FOR SEQ ID NO:36:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

His Leu Leu Glu Ile Ser Tyr Glu Ala Leu Glu Asn Ala Gly Phe
5 10 15

Pro Leu Pro Ser Ile Ala Gly Thr Asn Met Gly Val Phe Val Gly
20 25 30

Gly Ser Asn Ser Glu Tyr Arg Ala His Ile Gly Asn Asp Thr Asp
35 40 45

Asn Leu Pro Met Phe Glu Ala Thr Gly Asn Ala Glu Ser Leu Leu
50 55 60

Ala Asn Arg Val Ser Tyr Val Tyr Asp Leu His Gly Ala Ser Leu
65 70 75

Thr Ile Gly Thr Ala Cys Ser Val Glu Phe Ser Ser Phe Gly Xaa
80 85 90

Arg Val Ser Gln Leu Ala Ala Gly Lys Ser Ser Thr Ala Ile Val
95 100 105

Ala Gly Ser Val Val Arg Ile Val Pro Ser Ser Thr Ile Ser Pro
110 115 120

Ser Thr Met Lys Leu Leu Ser Pro Glu Gly Arg Cys Tyr Ala Phe
125 130 135

Asp Asp Arg Ala Thr Ser Gly Phe Gly Arg Gly Glu Gly Ser Ala
140 145 150

Cys Ile Ile Leu Glu Thr Leu Glu Ala Ala Leu Arg Asp Asn Asp
155 160 165

Pro Ile Arg Ser Val Ile Arg Asn Ser Gly Val Asn Gln Asp Gly
170 175 180

Lys Thr Ala Gly Ile Thr Met Pro Asn Gly Glu Ala Gln Ala Ser
185 190 195

Leu Ile Gln Ser Val Tyr Arg Thr Ala Gly Leu Asp Pro Leu Gln
200 205 210

Thr Asp Tyr Val Glu
215

(2) INFORMATION FOR SEQ ID NO:37
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:680
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AACTGTTAGA GGTCAGTTAC GAGGCGTTTG AGAATGCGGG CATATCATTA 50
TCGAGTGTTG CAGGTACCGA CGTTGGGGTA TTCATCAGTG CCAGCACCAA 100
TGATTACCGT TTCGTTTTCC ACAACGACCT CGACACATTG CCAATGTTTG 150
AATCCACTGG GAGTGAATTA TCGATCATGT CCAATCGTAT CTCCTATACT 200
TTCAATCTTA GAGGTCCAAG TATGACGATT GATACTCCCT GTTCCTCAAG 250
TTTGATCGCA CTCCATACAG CATTCAGAAG TCTACAGGTC GGAGAAAGCT 300
CTTGCGCCAT TGTCGGTGGA TCTAACCTCC ACATCACTCC AGATTCCTAC 350
ATTTCATTCT CGACGATGAG GTAAGCACTA TCGTTTGCGA ATTACCTATC 400
TTTGATTACG AGTGACTAAG TTGTACAGGC TCCTGTCGCC CCATGGACGA 450
TCGTGCAGTC AATGGGTTTG GGCGCGGAGA GGGCACAAGT TGCATAATAC 500
TGAAGCCTTT AGATGCCGCA TTGAAAGACC ACGATCCCAT AAGGGCAGTT 550
ATTCGCAATA CGGGCACTAA TCAAGATGGG AAGACGACAG GTATCACGAT 600
GCCGAATGGT GAAGCACAGG CCGCCTTAAT GCAATCAGTC TACGAGGCAG 650
CGGGCTTAGA TCCCCTTGAA ACAGACTATG 680

(2) INFORMATION FOR SEQ ID NO:38:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 209
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Leu Leu Glu Val Ser Tyr Glu Ala Phe Glu Asn Ala Gly Ile Ser
5 10 15

Leu Ser Ser Val Ala Gly Thr Asp Val Gly Val Phe Ile Ser Ala
20 25 30

Ser Thr Asn Asp Tyr Arg Phe Val Phe His Asn Asp Leu Asp Thr
35 40 45

Leu Pro Met Phe Glu Ser Thr Gly Ser Glu Leu Ser Ile Met Ser
50 55 60

Asn Arg Ile Ser Tyr Thr Phe Asn Leu Arg Gly Pro Ser Met Thr
65 70 75

Ile Asp Thr Pro Cys Ser Ser Ser Leu Ile Ala Leu His Thr Ala
80 85 90

Phe Arg Ser Leu Gln Val Gly Glu Ser Ser Cys Ala Ile Val Gly
95 100 105

Gly Ser Asn Leu His Ile Thr Pro Asp Ser Tyr Ile Ser Phe Ser
110 115 120

Thr Met Ser Cys Thr Gly Ser Cys Arg Pro Met Asp Asp Arg Ala
125 130 135

Val Asn Gly Phe Gly Arg Gly Glu Gly Thr Ser Cys Ile Ile Leu
140 145 150

Lys Pro Leu Asp Ala Ala Leu Lys Asp His Asp Pro Ile Arg Ala
155 160 165

Val Ile Arg Asn Thr Gly Thr Asn Gln Asp Gly Lys Thr Thr Gly
170 175 180

Ile Thr Met Pro Asn Gly Glu Ala Gln Ala Ala Leu Met Gln Ser
185 190 195

Val Tyr Glu Ala Ala Gly Leu Asp Pro Leu Glu Thr Asp Tyr
200 205

(2) INFORMATION FOR SEQ ID NO:39:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:691
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCATTTGCTG GAGGTGAGCT ATGAAGCGCT TGAAAATGCT GGCCTTTCTC 50
TTCCTTGCAT TGCCGGCACC AAAATGGGAG TCTTCGTTGG TGGAGGCAAT 100
GCAKAGTATC GATCGCATAT CGGCCAAGAT ATTGACAATC TGCCTATGTT 150
CGAGGCAACT GGTAACGCAG AGGCGCTATT GGCGAATAGA GTTTCTTATG 200
TATATGATCT TCGAGGACCG AGTCTAACCA CCGATACCGC CTGTTCCTCA 250
AGTCTCGCCG CTTTGAACAC GGCATTCTTA AGTCTACAGG CTGGCGAGTC 300
GTCTACAGCA CTGGTCGGTA GCTCAGTAAT TCGGCTTAGG CCTGAGTCAG 350
CCATCTCACT TTCCAGCATG CAGTAAGTCC TTCATGGTGC ACCTGCATAC 400
ATTGCTAATA AGTGCAGGCT TCTATCCCCA GATGGAAAAT CTTACGCGTT 450
CGATGAGAGA GCTACCAGTG GTTTTGGAAG GGGTGAGGGT TCGGGTTGCA 500
TAATACTAAA ACCCCTGGAC GCAGCCGTGA GAGACGGAGA CCCAATTAGA 550
GCAGTCATTT GTAACTCGGG TGTAAACCAA GACGGCAAGA CTGCTGGTAT 600
TACAATGCCT AATGGACACG CGCAAGCTTC TCTAATACGG TCTGTTTATC 650
AGTCTACAGG GATAGACCCT TTAATGACGG ACTATGTCGA A 691

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
His Leu Leu Glu Val Ser Tyr Glu Ala Leu Glu Asn Ala Gly Leu
                  5                  10                  15

Ser Leu Pro Cys Ile Ala Gly Thr Lys Met Gly Val Phe Val Gly
                 20                  25                  30

Gly Gly Asn Ala Xaa Tyr Arg Ser His Ile Gly Gln Asp Ile Asp
                 35                  40                  45

Asn Leu Pro Met Phe Glu Ala Thr Gly Asn Ala Glu Ala Leu Leu
                 50                  55                  60

Ala Asn Arg Val Ser Tyr Val Tyr Asp Leu Arg Gly Pro Ser Leu
                 65                  70                  75

Thr Thr Asp Thr Ala Cys Ser Ser Ser Leu Ala Ala Leu Asn Thr
                 80                  85                  90

Ala Phe Leu Ser Leu Gln Ala Gly Glu Ser Ser Thr Ala Leu Val
                 95                 100                 105

Gly Ser Ser Val Ile Arg Leu Arg Pro Glu Ser Ala Ile Ser Leu
                110                 115                 120

Ser Ser Met Gln Leu Leu Ser Pro Asp Gly Lys Ser Tyr Ala Phe
                125                 130                 135

Asp Glu Arg Ala Thr Ser Gly Phe Gly Arg Gly Glu Gly Ser Gly
                140                 145                 150

Cys Ile Ile Leu Lys Pro Leu Asp Ala Ala Val Arg Asp Gly Asp
                155                 160                 165

Pro Ile Arg Ala Val Ile Cys Asn Ser Gly Val Asn Gln Asp Gly
                170                 175                 180

Lys Thr Ala Gly Ile Thr Met Pro Asn Gly His Ala Gln Ala Ser
                185                 190                 195

Leu Ile Arg Ser Val Tyr Gln Ser Thr Gly Ile Asp Pro Leu Met
                200                 205                 210

Thr Asp Tyr Val Glu
                215
```

(2) INFORMATION FOR SEQ ID NO:41:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:637
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTGTTTCTT CAAACTAGCT GGCAATGCAT TGAAGATGCG GGATATAACC 50
CCACATCCTT TGCAGGTAGC AAGTGTGGCG TATTTGTCGG CTGCGAAACG 100
GGAGACTATG GAAAGATTGT GCAGCGATAT GAATTGAGCG CTCTCGGATT 150
GCTAGGCTCT TCTGCGGCAC TGCTCCCGGC AAGGATCTCC TATTTCCTCA 200
ACCTCCAGGG CCCTTGTATG GCGATCGACA CAGCCTGCTC TGCATCCCTA 250
GTTGCCATAG CCAACGCCTG CGACAGCCTG GTACTGGGTC ACTCCGATGC 300
AGCCTTGGCC GGAGGAGTCT ACGTCCTCTC CGGGCCGGAA ATGCACATTA 350
TGATGAGCAA AGCTGGTATC TTGTCACCCG ATGGCAGATG TTTCACCTTC 400
GATCGACGTG CTAACGGCTT TGTACCGGGC GAAGGTGTGG GCGTCGTGTT 450
ACTCAAACGC CTTGCCGATG CCGAAAAAGA CGGTGATAAT ATCTGTGGTG 500
TGATTCGAGG CTGGGGGGTG AATCAAGACG GCAAGACCAG TGGAATTACA 550
GCACCTAACG GACAGTCACA GCAACGATTG CAGAAAGAAG TCTACGAACG 600
GTTTCAGATT CAGCCAGCAG ACATTCAACT GGTTGAG 637

(2) INFORMATION FOR SEQ ID NO:42:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Leu Phe Leu Gln Thr Ser Trp Gln Cys Ile Glu Asp Ala Gly Tyr
5 10 15

Asn Pro Thr Ser Phe Ala Gly Ser Lys Cys Gly Val Phe Val Gly
20 25 30

Cys Glu Thr Gly Asp Tyr Gly Lys Ile Val Gln Arg Tyr Glu Leu
35 40 45

Ser Ala Leu Gly Leu Leu Gly Ser Ser Ala Ala Leu Leu Pro Ala
50 55 60

Arg Ile Ser Tyr Phe Leu Asn Leu Gln Gly Pro Cys Met Ala Ile
65 70 75

Asp Thr Ala Cys Ser Ala Ser Leu Val Ala Ile Ala Asn Ala Cys
80 85 90

Asp Ser Leu Val Leu Gly His Ser Asp Ala Ala Leu Ala Gly Gly
95 100 105

Val Tyr Val Leu Ser Gly Pro Glu Met His Ile Met Met Ser Lys
110 115 120

Ala Gly Ile Leu Ser Pro Asp Gly Arg Cys Phe Thr Phe Asp Arg
125 130 135

Arg Ala Asn Gly Phe Val Pro Gly Glu Gly Val Gly Val Val Leu
140 145 150

Leu Lys Arg Leu Ala Asp Ala Glu Lys Asp Gly Asp Asn Ile Cys
155 160 165

Gly Val Ile Arg Gly Trp Gly Val Asn Gln Asp Gly Lys Thr Ser
170 175 180

Gly Ile Thr Ala Pro Asn Gly Gln Ser Gln Gln Arg Leu Gln Lys
185 190 195

Glu Val Tyr Glu Arg Phe Gln Ile Gln Pro Ala Asp Ile Gln Leu
200 205 210

Val Glu

(2) INFORMATION FOR SEQ ID NO:43:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:643
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GATGATGATA GAAGTCGCTT ACCAAGGACT TGAGAGTGCA GGGCTGTCTC 50
TTCAGGATGT TGCCGGATCG AGGACTGGAG TCTTCATTGG CCATTTCAGC 100
AGTGATTACC GAGACATGAT ATTCAGAGAT CCCGAGAGGG CACCGACCTA 150
CACTTTCAGT GGGGTTAGTA AGACGTCATT GGCGAATCGC ATCTCCTGGC 200
TGTTCGACCT GAAAGGCCCA AGTTTCAGCT TGGACACAGC CTGCTCGTCG 250
AGTCTGGTCG CCCTGCATTT GGCTTGCCAA AGCTTACGCG CTGGAGAGTC 300
AGATATCGCC ATTGTCGGAG GGGTCAACCT TCTCTGGAAT CCGGAGTTGT 350
TCATGTATCT CTCCAATCAG CACTTTCTCT CGCCAGATGG GAAATGTAAA 400
AGCTTTGACG AATCCGGCGA TGGCTATGGT CGTGGCGAAG GCATTGCCGC 450
TCTTGTACTA AGAAGAGTCG ACGACGCGAT TGCGGCCCGG GACCCTATTC 500
GTGCCATCAT TCGCGGTACT GGGAGTAATC AGGACGGACA CACCAAAGGC 550
TTCACCCTCC CCAGCGCAGA AGCCCAGGCG AGGTTGATTA GAGATACGTA 600
CTCTGCCGCG GGGCTAGGTT TTAGAGACAC GCGATACGTA GAA 643

(2) INFORMATION FOR SEQ ID NO:44:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```
Met Met Ile Glu Val Ala Tyr Gln Gly Leu Glu Ser Ala Gly Leu
                  5                  10                  15

Ser Leu Gln Asp Val Ala Gly Ser Arg Thr Gly Val Phe Ile Gly
                 20                  25                  30

His Phe Ser Ser Asp Tyr Arg Asp Met Ile Phe Arg Asp Pro Glu
                 35                  40                  45

Arg Ala Pro Thr Tyr Thr Phe Ser Gly Val Ser Lys Thr Ser Leu
                 50                  55                  60

Ala Asn Arg Ile Ser Trp Leu Phe Asp Leu Lys Gly Pro Ser Phe
                 65                  70                  75

Ser Leu Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Leu
                 80                  85                  90

Ala Cys Gln Ser Leu Arg Ala Gly Glu Ser Asp Ile Ala Ile Val
                 95                 100                 105

Gly Gly Val Asn Leu Leu Trp Asn Pro Glu Leu Phe Met Tyr Leu
                110                 115                 120

Ser Asn Gln His Phe Leu Ser Pro Asp Gly Lys Cys Lys Ser Phe
                125                 130                 135

Asp Glu Ser Gly Asp Gly Tyr Gly Arg Gly Glu Gly Ile Ala Ala
                140                 145                 150

Leu Val Leu Arg Arg Val Asp Asp Ala Ile Ala Ala Arg Asp Pro
                155                 160                 165

Ile Arg Ala Ile Ile Arg Gly Thr Gly Ser Asn Gln Asp Gly His
                170                 175                 180

Thr Lys Gly Phe Thr Leu Pro Ser Ala Glu Ala Gln Ala Arg Leu
                185                 190                 195

Ile Arg Asp Thr Tyr Ser Ala Ala Gly Leu Gly Phe Arg Asp Thr
                200                 205                 210

Arg Tyr Val Glu
```

(2) INFORMATION FOR SEQ ID NO:45:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:655
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```
RGTCCTTATG GAGACCGTCT ACGAGGCAAT TGAGTCTGCG GGTATGACTT 50
TGAAGGGGCT GCAAGGCAGC GACACAAGTG TGTATGCCGG CGTCATGTGT 100
GGCGACTACG AGGCCATACA GCTCCGCGAT CTGGACGCGG CCCCGACTTA 150
TTTCGCAGTG GGAACCTCGC GAGCTATCCT CTCCAATCGA ATCTCGTATT 200
TCTTCAACTG GCACGGCGCG TCCATCACCA TGGACACGGC ATGTTCCTCT 250
AGTCTGGTCG CCATTCACTT GGCCGTTCAG RCGCTTCGGG CAAATGAATC 300
ACGRATGGCC GTGGCGTGTG GGTCGAACCT CATTCTCGGA CCCGAGAGTT 350
ACATTATTGA AAGCAAGGTG AAGATGCTGT CCCCGGACGG TCTCAGCCGA 400
ATGTGGGATA AAGACGCCAA CGGCTATGCG CGTGGAGATG GCGTTGCGGC 450
CGTTGTTTTG AAGACTCTCA GCGCCGCGCT GGCGGACGGA GACCACATTG 500
AATGTCTCAT ACGGGAGACG GGACTCAACC AGGACGGTGC GACAGCCGGT 550
CTCACCATGC CTAGCGCCAC TGCGCAGCGA GCTCTTATTC ACAGTACGTA 600
CACCAAGGCA GGTCTTGATC TCACTGCCCA GGCAGACCGT CCCCAGTATT 650
TCGAG                                                  655
```

(2) INFORMATION FOR SEQ ID NO:46:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
Val Leu Met Glu Thr Val Tyr Glu Ala Ile Glu Ser Ala Gly Met
                  5                  10                  15

Thr Leu Lys Gly Leu Gln Gly Ser Asp Thr Ser Val Tyr Ala Gly
                 20                  25                  30

Val Met Cys Gly Asp Tyr Glu Ala Ile Gln Leu Arg Asp Leu Asp
                 35                  40                  45

Ala Ala Pro Thr Tyr Phe Ala Val Gly Thr Ser Arg Ala Ile Leu
                 50                  55                  60

Ser Asn Arg Ile Ser Tyr Phe Phe Asn Trp His Gly Ala Ser Ile
                 65                  70                  75

Thr Met Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu
                 80                  85                  90

Ala Val Gln Xaa Leu Arg Ala Asn Glu Ser Arg Met Ala Val Ala
                 95                 100                 105

Cys Gly Ser Asn Leu Ile Leu Gly Pro Glu Ser Tyr Ile Ile Glu
                110                 115                 120

Ser Lys Val Lys Met Leu Ser Pro Asp Gly Leu Ser Arg Met Trp
```

125 130 135

Asp Lys Asp Ala Asn Gly Tyr Ala Arg Gly Asp Gly Val Ala Ala
140 145 150

Val Val Leu Lys Thr Leu Ser Ala Ala Leu Ala Asp Gly Asp His
155 160 165

Ile Glu Cys Leu Ile Arg Glu Thr Gly Leu Asn Gln Asp Gly Ala
170 175 180

Thr Ala Gly Leu Thr Met Pro Ser Ala Thr Ala Gln Arg Ala Leu
185 190 195

Ile His Ser Thr Tyr Thr Lys Ala Gly Leu Asp Leu Thr Ala Gln
200 205 210

Ala Asp Arg Pro Gln Tyr Phe Glu
215

(2) INFORMATION FOR SEQ ID NO:47:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:754
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AGGTCTGTTG GAGACGGTTT ATCGCGCCTT TGAAAACGGT AAGGCCACCC 50
TGGGAATAAA CCGGCTTCTC GTCCTGACGG CTTACTCTAT GCTAGCTGGT 100
ATACCCATGG AGCAGGTCCT CGGGTCGAAG ACATCCGTTT ACGTGGGATG 150
TTTCACCCGC GAGTTCGAGC AGTTGCTCGC GAGGGACCCC GAGATGAATC 200
TGAAATACAT CGCTACGGGC ACCGGCACGG CGATGCTGTC GAATCGCCTC 250
TCCTGGTTCT ATGACTTGAA AGGCGCCAGT ATCACTCTTG ATACTGCCTG 300
TTCGTCCAGT CTCAATGCGT GCCATCTTGC TTGCGCAAGC TTACGTAATG 350
GAGAAGCCAA TATGGTAAGA CTCCAACTCA TCGCGGGACT GAACAATTGC 400
ATACTGATCC ATCAAAGGCC CTGGTAGGAG GCTGCAATCT TTTCTATAAC 450
CCGGAAACGA TCATCCCTCT GACAAATCTA GGCTTTCTTT CTCCGGATAA 500
CAAATGTTAT AGTTTTGACC ATCGTGCTAA CGGTTACTCT CGCGGCGAGG 550
GGTTTGGTAT TCTTGTATTG AAGAGACTGT CGGACGCTCT ACGCGATAAC 600
GACACTGTCC GTGCAGTGAT TCGGGCCTCT TCGTCTAACC AGGATGGCAA 650
GTCTCCCGGT ATCACACAGC CTACCAAACA AGCGCAAATA CAACTGATCA 700
AAGACACTTA CGCGGCTGCC GGGCTGGACT ATACGCAAAC CCGCTACTTC 750
GANA 754

(2) INFORMATION FOR SEQ ID NO:48:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Gly Leu Leu Glu Thr Val Tyr Arg Ala Phe Glu Asn Ala Gly Ile
5 10 15

Pro Met Glu Gln Val Leu Gly Ser Lys Thr Ser Val Tyr Val Gly
20 25 30

Cys Phe Thr Arg Glu Phe Glu Gln Leu Leu Ala Arg Asp Pro Glu
35 40 45

Met Asn Leu Lys Tyr Ile Ala Thr Gly Thr Gly Thr Ala Met Leu
50 55 60

Ser Asn Arg Leu Ser Trp Phe Tyr Asp Leu Lys Gly Ala Ser Ile
65 70 75

Thr Leu Asp Thr Ala Cys Ser Ser Ser Leu Asn Ala Cys His Leu
80 85 90

Ala Cys Ala Ser Leu Arg Asn Gly Glu Ala Asn Met Ala Leu Val
95 100 105

Gly Gly Cys Asn Leu Phe Tyr Asn Pro Glu Thr Ile Ile Pro Leu
110 115 120

Thr Asn Leu Gly Phe Leu Ser Pro Asp Asn Lys Cys Tyr Ser Phe
125 130 135

Asp His Arg Ala Asn Gly Tyr Ser Arg Gly Glu Gly Phe Gly Ile
140 145 150

Leu Val Leu Lys Arg Leu Ser Asp Ala Leu Arg Asp Asn Asp Thr
155 160 165

Val Arg Ala Val Ile Arg Ala Ser Ser Ser Asn Gln Asp Gly Lys
170 175 180

Ser Pro Gly Ile Thr Gln Pro Thr Lys Gln Ala Gln Ile Gln Leu
185 190 195

Ile Lys Asp Thr Tyr Ala Ala Ala Gly Leu Asp Tyr Thr Gln Thr
200 205 210

Arg Tyr Phe Xaa

(2) INFORMATION FOR SEQ ID NO:49:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:722
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```
CTTGTTACTC GAGACTGTCT ACGAATCTCT CGAGTCGGCT GGTCAGACAA 50
TCGAAGGCTT GCAAGGATCG CAAACCGCAG TGTATATTGG TGTAATGTGC 100
GATGATTACG CCGAGCTCGT GTATCATGAT ACAGAGTCAA TCCCGACCTA 150
TGCTGCAACT GGTAGTGCAC GCAGCATGAT GTCGAACCGA ATCTCTTACT 200
TCTTTGACTG GAAGGGGCCG TCAATGACCA TTGATACTGC CTGTTCCTCT 250
AGTCTTGTCG CTGTCCACCA GGCCGTTCAA GTTCTCAGGA GCGGAGAATC 300
CCGCGTCGCA GTGGCTGCTG GGGCAAATCT CATCTTCGGA CCCAGTAAGT 350
CTTCCTAAAA TATGAGTAGG CTCCAGTCAT TGTGATTGCT AATCACTTCA 400
ACCATTTACA GAGATGTACA TTGCTGAGAG CAACCTCAAT ATGTTGTCCC 450
CAACTGGSCG STCCCGAATG TGGGACGCTA ACSCGGATGG CTATGCACGA 500
GGAGAGGGTA TTGCATCTGT CGTACTCAAA ACTCTTAGCT CTGCTATAGC 550
AGATGGTGAT ACCATCGAAT GTTTGATCCG AGAAACCGGT GTCAACCAGG 600
ATGGCCGCAC CACTGGTATC ACTATGCCAA GCTCCGCAGC CCAAGCCAGT 650
TTGATCCGTC AGACTTACGC CAGAGCTGGT TTGGACCTGG CGAAGCAAGC 700
TGATCGGCCT CAATTCTTTG AG                                722
```

(2) INFORMATION FOR SEQ ID NO:50:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

```
Leu Leu Leu Glu Thr Val Tyr Glu Ser Leu Glu Ser Ala Gly Gln
                  5                  10                  15

Thr Ile Glu Gly Leu Gln Gly Ser Gln Thr Ala Val Tyr Ile Gly
                 20                  25                  30

Val Met Cys Asp Asp Tyr Ala Glu Leu Val Tyr His Asp Thr Glu
                 35                  40                  45

Ser Ile Pro Thr Tyr Ala Ala Thr Gly Ser Ala Arg Ser Met Met
                 50                  55                  60

Ser Asn Arg Ile Ser Tyr Phe Phe Asp Trp Lys Gly Pro Ser Met
                 65                  70                  75

Thr Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Gln
                 80                  85                  90

Ala Val Gln Val Leu Arg Ser Gly Glu Ser Arg Val Ala Val Ala
                 95                 100                 105

Ala Gly Ala Asn Leu Ile Phe Gly Pro Lys Met Tyr Ile Ala Glu
                110                 115                 120
```

Ser Asn Leu Asn Met Leu Ser Pro Thr Gly Arg Ser Arg Met Trp
125 130 135

Asp Ala Asn Xaa Asp Gly Tyr Ala Arg Gly Glu Gly Ile Ala Ser
140 145 150

Val Val Leu Lys Thr Leu Ser Ser Ala Ile Ala Asp Gly Asp Thr
155 160 165

Ile Glu Cys Leu Ile Arg Glu Thr Gly Val Asn Gln Asp Gly Arg
170 175 180

Thr Thr Gly Ile Thr Met Pro Ser Ser Ala Ala Gln Ala Ser Leu
185 190 195

Ile Arg Gln Thr Tyr Ala Arg Ala Gly Leu Asp Leu Ala Lys Gln
200 205 210

Ala Asp Arg Pro Gln Phe Phe Glu
215

(2) INFORMATION FOR SEQ ID NO:51:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:703
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AATATTACTT GAGACGATCT ACGAAGGACT TGAGTCCGCC GGACTTACCA 50
TAAAGGGGCT GCAAGGTTCC CAAACAGCTG TGTACGTCGG TCTCATGGCT 100
GGAGACTACT ATGACATCCA GATGCGCGAC ATAGAGACTT TGCCTCGATA 150
TGCTGCTACC GGGACTGCTC GTAGCATTAT GAGCAACCGA GTCTCTTATT 200
TCTTTGATTG GAAAGGTCCG TCCATGACAA TTGATACGGC CTGCTCTTCT 250
TCCCTCGTTG CCGTTCATCA GGCTGTCGAG ATTCTCCGGA GAGGTGATGT 300
TACCATGGCT GTGGCTGCCG GCGCCAACCT GATCTATGGT CCTGAGGCTT 350
ATATATCCGA GTCGAATCTG AACATGCTGT CGCCGAGCGG AAGATCGCGC 400
ATGTGGGATT CAAGTGCGGA CGGATACGGC CGCGGAGAAG GGTTTGCGGC 450
AGTGATGTTG AAGACCCTGA GCGCTGCAAT TCGTGATGGA GATCATATCG 500
AGTGCATTAT CCGGGAGACA GGAATTAACC AGGATGGCAG AACAGCCGGA 550
ATTACCATGC CAAGTGCTGT CAGCCAGACT CGATTGATCA AAGACACATA 600
TGCTCGAGCT GGACTCGATT GCAGGAAAGA AGCGGAGAGA TGCCAGTACT 650
TTGAAGGTAA GCGAATAACT TTTCTTGATA AACGCACTTA CTAAGATCTT 700
TAA 703

(2) INFORMATION FOR SEQ ID NO:52:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 234
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Ile Leu Leu Glu Thr Ile Tyr Glu Gly Leu Glu Ser Ala Gly Leu
5 10 15

Thr Ile Lys Gly Leu Gln Gly Ser Gln Thr Ala Val Tyr Val Gly
20 25 30

Leu Met Ala Gly Asp Tyr Tyr Asp Ile Gln Met Arg Asp Ile Glu
35 40 45

Thr Leu Pro Arg Tyr Ala Ala Thr Gly Thr Ala Arg Ser Ile Met
50 55 60

Ser Asn Arg Val Ser Tyr Phe Phe Asp Trp Lys Gly Pro Ser Met
65 70 75

Thr Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Gln
80 85 90

Ala Val Glu Ile Leu Arg Arg Gly Asp Val Thr Met Ala Val Ala
65 70 75

Ala Gly Ala Asn Leu Ile Tyr Gly Pro Glu Ala Tyr Ile Ser Glu
110 115 120

Ser Asn Leu Asn Met Leu Ser Pro Ser Gly Arg Ser Arg Met Trp
125 130 135

Asp Ser Ser Ala Asp Gly Tyr Gly Arg Gly Glu Gly Phe Ala Ala
140 145 150

Val Met Leu Lys Thr Leu Ser Ala Ala Ile Arg Asp Gly Asp His
155 160 165

Ile Glu Cys Ile Ile Arg Glu Thr Gly Ile Asn Gln Asp Gly Arg
170 175 180

Thr Ala Gly Ile Thr Met Pro Ser Ala Val Ser Gln Thr Arg Leu
185 190 195

Ile Lys Asp Thr Tyr Ala Arg Ala Gly Leu Asp Cys Arg Lys Glu
200 205 210

Ala Glu Arg Cys Gln Tyr Phe Glu Gly Lys Arg Ile Thr Phe Leu
215 220 225

Asp Lys Arg Thr Tyr Xaa Asp Leu Xaa
230

(2) INFORMATION FOR SEQ ID NO:53:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:643
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```
GCTGTTGCTG GAGGTAAGTT GGGAAGCTTT AGAAAATGCT GGCAAAGCAC 50
CTGAAAAGCT AGCAGGAAGC AATACAGGTG TATTTGTTGG CATTAGCAAC 100
TTTGATTATT CACAGTTGCA AATTAATCAA ACCGCTCAAC TAGATGCCTA 150
TACAGGCACT GGCAATGCTT TTAGCATCGC AGCTAACCGT CTTTCCTATT 200
TTCTAGACTT GCACGGACCT AGCTGGGCAG TAGACACAGC CTGTTCATCA 250
TCTCTAGTAG CAGTCCATCA AGCTTGCCAA AGTCTGCGTC AAGGAGAATG 300
CGAACTAGCC CTCGCTGGTG GTGTAAATCT GATTCTCACC CCACAATTAA 350
CCATCACTTT TTCCCAAGCT GGGATGATGG CTGCTGATGG TCGTTGCAAA 400
ACCTTTGATG CTGATGCTGA TGGTTACGTG CGGGGCGAAG GTTGTGGTGT 450
TGTAATTCTC AAGCGTTTGG CCAACGCTCA ACGAGATGGA GACAATATTT 500
TGGCAGTTAT TAAAGGTTCG GCAGTTAACC AAGATGGTCG CAGCAACGGA 550
TTGACAGCAC CCAACGGTCA TGCCCAACAA GCAGTTATTC GCCAAGCATT 600
ACAAAATGCC AATGTTGCAG CTGCCGAGAT TAGCTATGTA GAA         643
```

(2) INFORMATION FOR SEQ ID NO:54:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```
Leu Leu Leu Glu Val Ser Trp Glu Ala Leu Glu Asn Ala Gly Lys
 5                  10                  15

Ala Pro Glu Lys Leu Ala Gly Ser Asn Thr Gly Val Phe Val Gly

                20                  25                  30

Ile Ser Asn Phe Asp Tyr Ser Gln Leu Gln Ile Asn Gln Thr Ala
            35                  40                  45

Gln Leu Asp Ala Tyr Thr Gly Thr Gly Asn Ala Phe Ser Ile Ala
        50                  55                  60

Ala Asn Arg Leu Ser Tyr Phe Leu Asp Leu His Gly Pro Ser Trp
    65                  70                  75

Ala Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Gln
                80                  85                  90

Ala Cys Gln Ser Leu Arg Gln Gly Glu Cys Glu Leu Ala Leu Ala
            95                  100                 105
```

Gly Gly Val Asn Leu Ile Leu Thr Pro Gln Leu Thr Ile Thr Phe
110 115 120

Ser Gln Ala Gly Met Met Ala Ala Asp Gly Arg Cys Lys Thr Phe
125 130 135

Asp Ala Asp Ala Asp Gly Tyr Val Arg Gly Glu Gly Cys Gly Val
140 145 150

Val Ile Leu Lys Arg Leu Ala Asn Ala Gln Arg Asp Gly Asp Asn
155 160 165

Ile Leu Ala Val Ile Lys Gly Ser Ala Val Asn Gln Asp Gly Arg
170 175 180

Ser Asn Gly Leu Thr Ala Pro Asn Gly His Ala Gln Gln Ala Val
185 190 195

Ile Arg Gln Ala Leu Gln Asn Ala Asn Val Ala Ala Ala Glu Ile
200 205 210

Ser Tyr Val Glu

(2) INFORMATION FOR SEQ ID NO:55:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:655
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TCTTTTTTTG GAGTGTGCTT GGGAAGCGCT GGAAAATGCT GGTTATGACC 50
CGAAAACAGA CAAAAATCTA ATTGGCGTTT ATGCAGGGGG GAATCTAAGT 100
ACCTACTTAC TTAACAATCT CGCCTCACAC CCTGAACTCA TTAAAGCGCT 150
GGAGTCACAA ATTACAATTG CTAATGATAA GGACTTTATA TGCACACGAG 200
TTTCTTACAA ATTAAACCTG AAAGGGCCGA GTATTAGTGT CGGCACGGCC 250
TGCTCTACGT CATTAGTAGC AGTTCACTTG GCATGTCGAG GATTGCTAAG 300
TTACCAGTGT GATATGGCAC TGGCTGGCGG TATTGCGATA CAAGTTCCAC 350
AAAAACAAGG TTATTTCTAT CAAGAAGGTG GCATGGCCTC TCCTGATGGC 400
CACTGTCGGG CCTTTGATGC TAAAGCACAA GGTAGCCCTT TTGGCAAAGG 450
AGCAGGTATT GTCGTGCTGA AAAGATTGGA AGATGCTGTA GCTGATGGAG 500
ACTGCATTTA TGCGGTTATC AAAGGTTCAG CCATCAATAA CGACGGTTCC 550
GAGAAGGTGA GTTACACCGC ACCCAGTGTA ACAGGCCAAG CAGAAGTGAT 600
TGCCGAGGCT CAGGCGATCG CTAACTTTGA TTCTGAAACA ATCACCTACA 650
TTGAA 655

(2) INFORMATION FOR SEQ ID NO:56:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 217
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

```
Leu Phe Leu Glu Cys Ala Trp Glu Ala Leu Glu Asn Ala Gly Tyr
                  5                  10                  15

Asp Pro Lys Thr Asp Lys Asn Leu Ile Gly Val Tyr Ala Gly Gly
                 20                  25                  30

Asn Leu Ser Thr Tyr Leu Leu Asn Asn Leu Ala Ser His Pro Glu
                 35                  40                  45

Leu Ile Lys Ala Leu Glu Ser Gln Ile Thr Ile Ala Asn Asp Lys
                 50                  55                  60

Asp Phe Ile Cys Thr Arg Val Ser Tyr Lys Leu Asn Leu Lys Gly
                 65                  70                  75

Pro Ser Ile Ser Val Gly Thr Ala Cys Ser Thr Ser Leu Val Ala
                 80                  85                  90

Val His Leu Ala Cys Arg Gly Leu Leu Ser Tyr Gln Cys Asp Met
                 95                 100                 105

Ala Leu Ala Gly Gly Ile Ala Ile Gln Val Pro Gln Lys Gln Gly
                110                 115                 120

Tyr Phe Tyr Gln Glu Gly Gly Met Ala Ser Pro Asp Gly His Cys
                125                 130                 135

Arg Ala Phe Asp Ala Lys Ala Gln Gly Ser Pro Phe Gly Lys Gly
                140                 145                 150

Ala Gly Ile Val Val Leu Lys Arg Leu Glu Asp Ala Val Ala Asp
                155                 160                 165

Gly Asp Cys Ile Tyr Ala Val Ile Lys Gly Ser Ala Ile Asn Asn
                170                 175                 180

Asp Gly Ser Glu Lys Val Ser Tyr Thr Ala Pro Ser Val Thr Gly
                185                 190                 195

Gln Ala Glu Val Ile Ala Glu Ala Gln Ala Ile Ala Asn Phe Asp
                200                 205                 210


Ser Glu Thr Ile Thr Tyr Ile
                215
```

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:765
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ATTGCTGCTT GAAAACGTCT ATGAAGCTCT TGAAAACGGT GAGCGGTTCT 50
TCAAGAGAAT ATTGATGCAT CAATATGCTA ACTTGATGTC AATCATCAGC 100
TGGTATTCCT CTGAGCGAGT CCGTCTCTTC TAACACCTCC GTTTATGTTG 150
GCTCATTCGG TGATGACTAT AAGACGATTC TCAATACCGA TTTTGAGAGT 200
TGGGTCAAGT ACAAAGGCAC CGGTGTCTAT AACTCGATTC TGGCCAATCG 250
AATCAGCTGG TTCTACGACT TTAAAGGAGC CAGCGTCACG CTAGATACCG 300
CATGCTCGAG TAGCTTGGTA GCCGTGCATA TGGCTTGCCA GGATTTGAGG 350
TTGGGAGAGT CTAGAATGGT CAGTGTATTT CTCTATTGAA AAGTACTAGA 400
GGATTCTAAT TGACGTATTT GGATACCAGT CCGTTGTCGG CGGTGTCAAC 450
ATCATTGGCC ATCCGTTGCT CGTCCACGAT CTAAGCAAGC TCGGAGCGCT 500
CTCTCCTGAT GGCGTGTGCT ACACTTTCGA TGAACGGGCC AATGGATATT 550
CCCGGGGAGA AGGTGTCGGC ACCATCGTTC TCAAACGGCT CTCTGACGCA 600
ATCGAAGATG GTGATACCAT TCGCGCTATC ATCCGTGCAA GCGGGTGCAA 650
TCAAGACGGT AAAACAGCAG GTATATTTGT CCCTTCAGTC CAAGCCCAGG 700
AGCGACTTAT CCGGGATACC TATGAGAAGG CTGGGCTTGA CCGGACACGC 750
ACGACATATT TGGAA 765

(2) INFORMATION FOR SEQ ID NO:58:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 216
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Leu Leu Leu Glu Asn Val Tyr Glu Ala Leu Glu Asn Ala Gly Ile
5 10 15

Pro Leu Ser Glu Ser Val Ser Ser Asn Thr Ser Val Tyr Val Gly
20 25 30

Ser Phe Gly Asp Asp Tyr Lys Thr Ile Leu Asn Thr Asp Phe Glu
35 40 45

Ser Trp Val Lys Tyr Lys Gly Thr Gly Val Tyr Asn Ser Ile Leu
50 55 60

Ala Asn Arg Ile Ser Trp Phe Tyr Asp Phe Lys Gly Ala Ser Val
65 70 75

Thr Leu Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Met
80 85 90

Ala Cys Gln Asp Leu Arg Leu Gly Glu Ser Arg Met Val Ser Ser
95 100 105

Val Val Gly Gly Val Asn Ile Ile Gly His Pro Leu Leu Val His
110 115 120

Asp Leu Ser Lys Leu Gly Ala Leu Ser Pro Asp Gly Val Cys Tyr
125 130 135

Thr Phe Asp Glu Arg Ala Asn Gly Tyr Ser Arg Gly Glu Gly Val
140 145 150

Gly Thr Ile Val Leu Lys Arg Leu Ser Asp Ala Ile Glu Asp Gly
155 160 165

Asp Thr Ile Arg Ala Ile Ile Arg Ala Ser Gly Cys Asn Gln Asp
170 175 180

Gly Lys Thr Ala Gly Ile Phe Val Pro Ser Val Gln Ala Gln Glu
185 190 195

Arg Leu Ile Arg Asp Thr Tyr Glu Lys Ala Gly Leu Asp Arg Thr
200 205 210

Arg Thr Thr Tyr Leu Glu
215

(2) INFORMATION FOR SEQ ID NO:59:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:709
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TAAGTTACTG GAAACAGCAT ATACTGCGTT TGAGAACGGT GAGTACGCCT 50
TGCGTCGTAT CCCCTCCCCC CTCATGGAAG ATCTCAATCT GATCTCGTGA 100
AACAGCCGGC ATCGGGTTAG AAGCGGCACG AGGATCAAAC ACTTCAGTAC 150
ATATAGGTTG TTTTAATATC GACTATACAA GCAACCATAG TAGAGATCCA 200
GAGCAGATGC ACAAATATAC GGGGACTGGA GGAGCACCTT CCATGCTGTC 250
GAACAGACTG AGTTGGTTTT TCGATCTGAG AGGACCGAGC TTGACCTTGG 300
ACACGGCATG CTCTAGTAGC ATGGTTGCGC TTGATTTAGC ATGCCAGACT 350
TTGCAAAGTG GACAATCTGA CATGGGTCTT GTCGGGGGTT GTAATCTCAT 400
CTACAGCGTC GACATGACCA TGGCTCTATC CAAGCTTGGA TTTCTCTCCC 450
ATAACAGTCG GTGCTACAGT TTTGACCATC GAGCGGATGG GTACGCCAGA 500
GGTGAAGGCT TTGGAGTTTT AATTCTCAAA CGTGTCGAAG ACGCCATACG 550
AGATGGGGAT ACTATACGAG GAGTCATTCG ATTAACAAGC TCCAATCAAG 600
ACGGCCATAC TCCGGGAATA ACAATGCCCA GCAGAGACGC CCAAGCAAGT 650
TTGATTAGAA AGACATACCA ACAAGCTGGA TTAGATATGC AGATGACAGG 700
CTACTTTGA 709

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Lys Leu Leu Glu Thr Ala Tyr Thr Ala Phe Glu Asn Ala Gly Ile
5 10 15

Gly Leu Glu Ala Ala Arg Gly Ser Asn Thr Ser Val His Ile Gly
20 25 30

Cys Phe Asn Ile Asp Tyr Thr Ser Asn His Ser Arg Asp Pro Glu
35 40 45

Gln Met His Lys Tyr Thr Gly Thr Gly Gly Ala Pro Ser Met Leu
50 55 60

Ser Asn Arg Leu Ser Trp Phe Phe Asp Leu Arg Gly Pro Ser Leu
65 70 75

Thr Leu Asp Thr Ala Cys Ser Ser Ser Met Val Ala Leu Asp Leu
80 85 90

Ala Cys Gln Thr Leu Gln Ser Gly Gln Ser Asp Met Gly Leu Val
95 100 105

Gly Gly Cys Asn Leu Ile Tyr Ser Val Asp Met Thr Met Ala Leu
110 115 120

Ser Lys Leu Gly Phe Leu Ser His Asn Ser Arg Cys Tyr Ser Phe
125 130 135

Asp His Arg Ala Asp Gly Tyr Ala Arg Gly Glu Gly Phe Gly Val
140 145 150

Leu Ile Leu Lys Arg Val Glu Asp Ala Ile Arg Asp Gly Asp Thr
155 160 165

Ile Arg Gly Val Ile Arg Leu Thr Ser Ser Asn Gln Asp Gly His
170 175 180

Thr Pro Gly Ile Thr Met Pro Ser Arg Asp Ala Gln Ala Ser Leu
185 190 195

Ile Arg Lys Thr Tyr Gln Gln Ala Gly Leu Asp Met Gln Met Thr
200 205 210

Gly Tyr Phe

(2) INFORMATION FOR SEQ ID NO:61:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:649
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

```
AATGTTGCTC GAGATCACCT ACGAAGCCCT GGAGAACGCT GGACTTCCTT 50
TGAGTAAGGT TGTCGGCTCT GATACAGCCT GCTTCATTGG TGGCTTTACA 100
CGAGATTATG ATGATTTGAC CACTTCGGAG CTCGCGAAGA CCCTACTCTA 150
CACAACTACC GGCAACGGCC TGACGATGAT GTCGAATCGC TTATCCTGGT 200
TCTACGACCT TCATGGCCCG TCGGTTTCGC TCGACACAGC ATGTTCTAGC 250
TCGCTGGTTG CACTAAACCT TGCATGCCAG ACAATCCGAG CATCGACGAA 300
TGACTCTCGA CAGGCGATAG TTGGAGGTGT CAATCTCATG CTGCTCCCTG 350
ATCAGATGAC CACGATTAAT CCTCTGCATT TCTTAAGTCC TGATAGCCAA 400
TGCTACTCGT TTGATGACCG TGCAAACGGT TACACCCGTG GAGAAGGTAT 450
TGGCATACTG GTGCTCAAGC ACATCAATGA TGCTATTCGA GATGGAGACT 500
GTATAAGGGC AGTAATCCGC GGCACTGGGG TCAACTCCGA TGGCAAGACC 550
CCTGGCATTA CCTTGCCAAG CACGGCTGCA CAAGCCTCTT TAATTCGCGC 600
AACGTACGCC TCGGCAGGGC TGGACCCAGC TCACACCGGC TACTTTGAA  649
```

(2) INFORMATION FOR SEQ ID NO:62:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 216
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```
Met Leu Leu Glu Ile Thr Tyr Glu Ala Leu Glu Asn Ala Gly Leu
 5                  10                  15

Pro Leu Ser Lys Val Val Gly Ser Asp Thr Ala Cys Phe Ile Gly
            20                  25                  30

Gly Phe Thr Arg Asp Tyr Asp Asp Leu Thr Thr Ser Glu Leu Ala
        35                  40                  45

Lys Thr Leu Leu Tyr Thr Thr Thr Gly Asn Gly Leu Thr Met Met
    50                  55                  60

Ser Asn Arg Leu Ser Trp Phe Tyr Asp Leu His Gly Pro Ser Val
65                  70                  75

Ser Leu Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu Asn Leu
            80                  85                  90

Ala Cys Gln Thr Ile Arg Ala Ser Thr Asn Asp Ser Arg Gln Ala
        95                  100                 105

Ile Val Gly Gly Val Asn Leu Met Leu Leu Pro Asp Gln Met Thr
    110                 115                 120
```

Thr Ile Asn Pro Leu His Phe Leu Ser Pro Asp Ser Gln Cys Tyr
125 130 135

Ser Phe Asp Asp Arg Ala Asn Gly Tyr Thr Arg Gly Glu Gly Ile
140 145 150

Gly Ile Leu Val Leu Lys His Ile Asn Asp Ala Ile Arg Asp Gly
155 160 165

Asp Cys Ile Arg Ala Val Ile Arg Gly Thr Gly Val Asn Ser Asp
170 175 180

Gly Lys Thr Pro Gly Ile Thr Leu Pro Ser Thr Ala Ala Gln Ala
185 190 195

Ser Leu Ile Arg Ala Thr Tyr Ala Ser Ala Gly Leu Asp Pro Ala
200 205 210

His Thr Gly Tyr Phe Glu
215

(2) INFORMATION FOR SEQ ID NO:63:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:747
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TATGCTACTT GAATGCACAT ACGAAGCGTT AGAGAATGGT CAGTGAGCTA 50
CGAGCCGATT TTCATATATC ATGGCTAACA AGTTGAAGCT GGCATACCTC 100
TAGATAAAGT AGTAGGAGAA CCCGTAGGGG TGTACGTCGG CTCAGCTAGT 150
TCCGATTACT CGGACATCGT GAACTCAGAC GGCGAGATGG TCTCCACTTA 200
CACGGCCACG GGGTTGGCCG CAACGATGAT GGCAAACCGC ATATCCTATT 250
TCTATGATCT CCGGGGGCCA AGCTTCACAT TGGACACGGC GTGTTCATCG 300
AGTTTGATGG CGTTACACCT AGCGTGCCAA AGTCTTCGAG TCGGTGAATC 350
GAAGCAAGCC ATTGTGGGCG GGGTCCACCT TGTACTGAGC CCGGATTGTA 400
TGACTTCGAT GAGTTTATTA GGGTAAGACC TTCAAAATCT CCATGCAGAA 450
TTTCTAAATC TAACCTACCA CCCTAGTTTG TTCTCTAATG ACGGCCGATC 500
CTACACTTAT GACCATCGAG GTACTGGTTA TGGGCGCGGC GAAGGTATTG 550
CTACCTTAGT AATAAAACCT CTTAAAGATG CGATGGAAGC CGGTGATAAC 600
ATCCGGGCCA TCATCCGCAA TAGTGGGGCA AATCAAGATG GTCGAACACC 650
AGGTGTGACT TTTCCAAGTC AAGATGCTCA GATAGATCTT ATGAGATCGG 700
TATATCGTTC CGCTGGACTT GATGTACTTG ATACCGGCTA CGTGGAA 747

(2) INFORMATION FOR SEQ ID NO:64:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Met Leu Leu Glu Cys Thr Tyr Glu Ala Leu Glu Asn Ala Gly Ile
5 10 15

Pro Leu Asp Lys Val Val Gly Glu Pro Val Gly Val Tyr Val Gly
20 25 30

Ser Ala Ser Ser Asp Tyr Ser Asp Ile Val Asn Ser Asp Gly Glu
35 40 45

Val Ser Thr Tyr Thr Ala Thr Gly Leu Ala Ala Thr Met Met
50 55 60

Ala Asn Arg Ile Ser Tyr Phe Tyr Asp Leu Arg Gly Pro Ser Phe
65 70 75

Thr Leu Asp Thr Ala Cys Ser Ser Ser Leu Met Ala Leu His Leu
80 85 90

Ala Cys Gln Ser Leu Arg Val Gly Glu Ser Lys Gln Ala Ile Val
95 100 105

Gly Gly Val His Leu Val Leu Ser Pro Asp Cys Met Thr Ser Met
110 115 120

Ser Leu Leu Gly Leu Phe Ser Asn Asp Gly Arg Ser Tyr Thr Tyr
125 130 135

Xaa His Arg Gly Thr Gly Tyr Gly Arg Gly Xaa Gly Ile Ala Thr
140 145 150

Leu Val Ile Lys Pro Leu Lys Asp Ala Met Glu Ala Gly Asp Asn
155 160 165

Ile Arg Ala Ile Ile Arg Asn Ser Gly Ala Asn Gln Asp Gly Arg
170 175 180

Thr Pro Gly Val Thr Phe Pro Ser Gln Asp Ala Gln Ile Asp Leu
185 190 195

Met Arg Ser Val Tyr Arg Ser Ala Gly Leu Asp Val Leu Asp Thr
200 205 210

Gly Tyr Val Glu

(2) INFORMATION FOR SEQ ID NO:65:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 643
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65

```
AATTCTACTT GAAGTCGCCT ATCAAGCAAT GGAGTCAAGC GGCTGCTTAC 50
GGAACCATCG ACGCGAAGCT GGGGATCCTG TGGGATGTTT TATTGGAGCT 100
AGCTTTGCCG AATATCTTGA CAACACCTGT TCTAATCCGC CAACCAGCTA 150
TACTTCCACT GGCACCATCA GAGCTTTCCA CTGCGGTAGA CTCAGTTATT 200
ACTTTGGATG GAGCGGTCCT GCCGAGGTCA TTGATACAGC TTGCTCCTCT 250
TCGTTGGTTG CTATCAATCG AGCTTGCAAG TCAGTGCAGG CGGGTGAATG 300
TACAATGGCT CTTACTGGTG GAGTGAACAT TATAACTGGT ATCCACAACT 350
TCTTAGATCT GGCAAAGGCT GGCTTYTTAA GCCCCACAGG CCAATGCAGA 400
CCCTTTGACC AGTCTGCAGA TGGGTATTGT CGCTCAGAAG GAGCAGGACT 450
TGTTGTACTA AAACTGTTAA GCCAAGCCAT AGCAGATGGA GATCAAATTT 500
TCGGAGTTAT TCCAAGTGTG TCCACCAACC AAGGCGGATT GTCATCTTCA 550
ATTACGATTC CTCATTCGCC TGCACAAAAA AAGTTGTATC AAACCGTGCT 600
TCGGCAAGCC GGCATGAAGC TAGAACAGGT TAGCTACGTA GAG         643
```

(2) INFORMATION FOR SEQ ID NO:66:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Ile Leu Leu Glu Val Ala Tyr Gln Ala Met Glu Ser Ser Gly Cys
                  5                  10                  15

Leu Arg Asn His Arg Arg Glu Ala Gly Asp Pro Val Gly Cys Phe
                 20                  25                  30

Ile Gly Ala Ser Phe Ala Glu Tyr Leu Asp Asn Thr Cys Ser Asn
                 35                  40                  45

Pro Pro Thr Ser Tyr Thr Ser Thr Gly Thr Ile Arg Ala Phe His
                 50                  55                  60

Cys Gly Arg Leu Ser Tyr Tyr Phe Gly Trp Ser Gly Pro Ala Glu
                 65                  70                  75

Val Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile Asn Arg
                 80                  85                  90

Ala Cys Lys Ser Val Gln Ala Gly Glu Cys Thr Met Ala Leu Thr
                 95                 100                 105

Gly Gly Val Asn Ile Ile Thr Gly Ile His Asn Phe Leu Asp Leu
                110                 115                 120

Ala Lys Ala Gly Phe Leu Ser Pro Thr Gly Gln Cys Arg Pro Phe
                125                 130                 135
```

Asp Gln Ser Ala Asp Gly Tyr Cys Arg Ser Glu Gly Ala Gly Leu
140 145 150

Val Val Leu Lys Leu Leu Ser Gln Ala Ile Ala Asp Gly Asp Gln
155 160 165

Ile Phe Gly Val Ile Pro Ser Val Ser Thr Asn Gln Gly Gly Leu
170 175 180

Ser Ser Ser Ile Thr Ile Pro His Ser Pro Ala Gln Lys Lys Leu
185 190 195

Tyr Gln Thr Val Leu Arg Gln Ala Gly Met Lys Leu Glu Gln Val
200 205 210

Ser Tyr Val Glu

(2) INFORMATION FOR SEQ ID NO:67:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:809
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AGGAAACTAC TAGAGGTCGT GTTTGAATGT TTTGAGAGTG CCGGTACACC 50
ACTTCACGCA GTTTCAGGAG CTAATATTGG CTGCTATGTT GGGAATTTTA 100
CGTTGGATTA TCTTGTCATG CAGTCTAAGG ATACAGACTC TTTTCATCGA 150
TATACTGCTC CAGGAATGGG ACCTACATTG TTAGCTAACC GCATAAGTCA 200
TGTTTTTAAT CTTCAAGGTC CAAGTGTTAT GCTTGATACA GCGTGTTCTT 250
CATCGATCTA CGCTCTTCAT GCAGCTTGTG TGGCCTTGAA TGCAGATGAG 300
TGCAATGCAG CAATTGTTGC TGGGGCAAAC CTAATCCAGT CACCTGAGTG 350
GCATCTTGCA GTCTCCAAAT CAGGTGTGAT TTCACAAACT TCCACGTGTC 400
ACACTTTCGA TGCTAGTGCG GATGGTTATG GGCGAGGCGA GGGCGTTGGG 450
GCCCTCTATC TCAAGCGTCT AAGTGACGCA ATCCGAGATC GAGATCCTAT 500
ACGGTCTGTT ATTCGTGGTA CAGCTGTTAA TAGGTTAGTA CATCCTCTTA 550
CCTTTCTTTC ATGGATTAGC GAGAATTAGG GTTCCAAATG TTTGAAAGCT 600
CGGGTTCTAA TATTCATTCA CTGGACTAGT AATGGCAAGA CAAACGGCAT 650
CAGTCAGCCT AGTGCTTTGG CACAGGAAGC TGTGATTAAA AAAGCTTATG 700
CAAAGGCGGG ATTACCTGTT ACCGAGACTG ACTATGTTGA GGTAAGTGAG 750
CTATGTTTAA ATCAGAAAAC GTCATGCCAT TATTTCTTAT CCTTCACTGA 800
NCTCTTACA 809

(2) INFORMATION FOR SEQ ID NO:68:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 237
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein

(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Arg Lys Leu Leu Glu Val Val Phe Glu Cys Phe Glu Ser Ala Gly
5 10 15

Thr Pro Leu His Ala Val Ser Gly Ala Asn Ile Gly Cys Tyr Val
20 25 30

Gly Asn Phe Thr Leu Asp Tyr Leu Val Met Gln Ser Lys Asp Thr
35 40 45

Asp Ser Phe His Arg Tyr Thr Ala Pro Gly Met Gly Pro Thr Leu
50 55 60

Leu Ala Asn Arg Ile Ser His Val Phe Asn Leu Gln Gly Pro Ser
65 70 75

Val Met Leu Asp Thr Ala Cys Ser Ser Ser Ile Tyr Ala Leu His
80 85 90

Ala Ala Cys Val Ala Leu Asn Ala Asp Glu Cys Asn Ala Ala Ile
95 100 105

Val Ala Gly Ala Asn Leu Ile Gln Ser Pro Glu Trp His Leu Ala
110 115 120

Val Ser Lys Ser Gly Val Ile Ser Gln Thr Ser Thr Cys His Thr
125 130 135

Phe Asp Ala Ser Ala Asp Gly Tyr Gly Arg Gly Glu Gly Val Gly
140 145 150

Ala Leu Tyr Leu Lys Arg Leu Ser Asp Ala Ile Arg Asp Arg Asp
155 160 165

Pro Ile Arg Ser Val Ile Arg Gly Thr Ala Val Asn Ser Asn Gly
170 175 180

Lys Thr Asn Gly Ile Ser Gln Pro Ser Ala Leu Ala Gln Glu Ala
185 190 195

Val Ile Lys Lys Ala Tyr Ala Lys Ala Gly Leu Pro Val Thr Glu
200 205 210

Thr Asp Tyr Val Glu Val Ser Glu Leu Cys Leu Asn Gln Lys Thr
215 220 225

Ser Cys His Tyr Phe Leu Ser Phe Thr Xaa Leu Leu
230 235

(2) INFORMATION FOR SEQ ID NO:69:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 658
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TTTGCTCCTT GAGACTGTCT ACGAAGCTCT GGAAGCAGGC GGTCACACGA 50
TTGAAGCGCT ACGAGGATCT GATACGTCTG TCTTTACAGG CACCATGGGC 100
GTCGACTACA ACGATACTGT TATACGGGAC CTGAACGTCA TCCCGACGTA 150
CTTTGCTACT GGAGTAAATC GAGCTATCAT CTCGAACCGA GTCTCATACT 200
TCTTTGACTG GCATGGGCCG AGCATGACCA TCGACACAGC CTGTTCATCC 250
AGTCTCGTCG CCGTGCACCA AGGAGTGAAA GCTCTTCGGA GTGGGGAGTC 300
GCGTACTGCC CTGGCATGTG GGACGCAGGT CATTCTAAAT CCCGAGATGT 350
ATGTTATTGA GAGCAAGCTG AAAATGCTTT CTCCTACGGG CCGCTCCCGC 400
ATGTGGGATG CGGACGCGGA TGGCTACGCT CGTGGGGAGG GCGTAGCGGC 450
TGTAGTGCTG AAACGGCTCA GTGACGCTAT TGCGGATGGA SATCGCATCG 500
AGTGCATCAT CCGTGAGACA GGGTCCAACC AAGACGGCCA TTCAAATGGT 550
ATCACGGTGC CGAGTACGGA GGCCCAAGCG GCCCTCATCC ACCAAACCTA 600
TGCCAGAGCT GGTCTAGACC CGGAAAATAA CCCTCACGAC CGCCCTCAGT 650
TCTTCGAA 658

(2) INFORMATION FOR SEQ ID NO:70:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Leu Leu Leu Glu Thr Val Tyr Glu Ala Leu Glu Ala Gly Gly His
5 10 15

Thr Ile Glu Ala Leu Arg Gly Ser Asp Thr Ser Val Phe Thr Gly
20 25 30

Thr Met Gly Val Asp Tyr Asn Asp Thr Val Ile Arg Asp Leu Asn
35 40 45

Val Ile Pro Thr Tyr Phe Ala Thr Gly Val Asn Arg Ala Ile Ile
50 55 60

Ser Asn Arg Val Ser Tyr Phe Phe Asp Trp His Gly Pro Ser Met
65 70 75

Thr Ile Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val His Gln
80 85 90

Gly Val Lys Ala Leu Arg Ser Gly Glu Ser Arg Thr Ala Leu Ala
95 100 105

Cys Gly Thr Gln Val Ile Leu Asn Pro Glu Met Tyr Val Ile Glu
110 115 120

Ser Lys Leu Lys Met Leu Ser Pro Thr Gly Arg Ser Arg Met Trp
125 130 135

Asp Ala Asp Ala Asp Gly Tyr Ala Arg Gly Glu Gly Val Ala Ala
140 145 150

Val Val Leu Lys Arg Leu Ser Asp Ala Ile Ala Asp Gly --- Arg
155 160 165

Ile Glu Cys Ile Ile Arg Glu Thr Gly Ser Asn Gln Asp Gly His
170 175 180

Ser Asn Gly Ile Thr Val Pro Ser Thr Glu Ala Gln Ala Ala Leu
185 190 195

Ile His Gln Thr Tyr Ala Arg Ala Gly Leu Asp Pro Glu Asn Asn
200 205 210

Pro His Asp Arg Pro Gln Phe Phe Glu
215

(2) INFORMATION FOR SEQ ID NO:71:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:753
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TGGGCTACTC GAGACTGCTT ACAAGGCGTT CGAAAACGGT GAGTCTTGAA 50
GCTGCACAGA TCAAGACAAG AACACTAAAT CTCTCAGCGG GCATACGCAT 100
AGAAGAAGCC GCTGGCTCTA GAACTTCAGT TCATATCGGG AGTTTCACTC 150
ATGATTGGAG AGACATCCTC CAAAGGGATC CACTAATGGA TGTTAGCTAC 200
ATAGCTACCG CAACCGAGGT TTCTATGCTA GCGAGTCGAC TCAGCTGGTT 250
TTATGATCTA AGTGGGCCYA GCATCTCCTT GGATACAGCG TGTTCGAGTA 300
GCTTAATGGC TTTACATCTC GCCTGCCAGA GTCTAAAGAG TCGAGAGGCC 350
GACATGGTAA GGCTATGCTA CTTTCTGGCT CACTCAAACT GTTTTCCATA 400
TCTGATGCTT GCACAGGGCC TTGTTGGGAG GGGCTAATCT TCTTTTGGAT 450
CCTGTAGGGG TTATTGGCAT AACAAATGTT GGCATGCTTT CGCCAGATGG 500
CATTAGTTAC AGCTTTGATC ATCGTGCAAA CGGGTATGCC CGAGGAGAAG 550
GGTTCGGAGT CGTTGTCATC AAACGCTTGG ACGATGCTCT CAGACATGGC 600
GATACTATTC GCGGTATCGT TCGTGCCACA GGATCGAATC AAGATGGAAG 650
AACTCCAGGG ATTACCCAAC CTGATGGAGC CGCGCAAGAA GAGCTCATCC 700
GAGACACTTA CAAAGCTGCT GGCTTAGATA TGAGGCTAGT AAGGTATTCT 750
TAA 753

(2) INFORMATION FOR SEQ ID NO:72:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Gly Leu Leu Glu Thr Ala Tyr Lys Ala Phe Glu Asn Ala Gly Ile
5 10 15

Arg Ile Glu Glu Ala Ala Gly Ser Arg Thr Ser Val His Ile Gly
20 25 30

Ser Phe Thr His Asp Trp Arg Asp Ile Leu Gln Arg Asp Pro Leu
35 40 45

Met Asp Val Ser Tyr Ile Ala Thr Ala Thr Glu Val Ser Met Leu
50 55 60

Ala Ser Arg Leu Ser Trp Phe Tyr Asp Leu Ser Gly Pro Ser Ile
65 70 75

Ser Leu Asp Thr Ala Cys Ser Ser Ser Leu Met Ala Leu His Leu
80 85 90

Ala Cys Gln Ser Leu Lys Ser Arg Glu Ala Asp Met Gly Leu Val
95 100 105

Gly Gly Ala Asn Leu Leu Leu Asp Pro Val Gly Val Ile Gly Ile
110 115 120

Thr Asn Val Gly Met Leu Ser Pro Asp Gly Ile Ser Tyr Ser Phe
125 130 135

Asp His Arg Ala Asn Gly Tyr Ala Arg Gly Glu Gly Phe Gly Val
140 145 150

Val Val Ile Lys Arg Leu Asp Asp Ala Leu Arg His Gly Asp Thr
155 160 165

Ile Arg Gly Ile Val Arg Ala Thr Gly Ser Asn Gln Asp Gly Arg
170 175 180

Thr Pro Gly Ile Thr Gln Pro Asp Gly Ala Ala Gln Glu Glu Leu
185 190 195

Ile Arg Asp Thr Tyr Lys Ala Ala Gly Leu Asp Met Arg Leu Val
200 205 210

Arg Tyr Ser

(2) INFORMATION FOR SEQ ID NO:73:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:753
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ATTGTTGCTC GAAGTAACCT ATGAAGCTTT AGAGAACGGT GGGTAGTTCC 50
AGGAAGCATT AATCAAGACA AAGCTATTGC TCACACTTTT CCAAAATAGC 100
CGGAATACCC TTGAACCAAA TTGTGGGCCA GGATGTTGGG GTTTTTGTTG 150
GCGGCTCAAT GTCCGACTAC CAGAACCTCC TCCACAAAGA CATCGCAAAT 200
GGTCCTATTT ACCAAGCCAC TGGCACTGCC ATGAGCTTCC TAGCCAACCG 250
AATATCTTAC ATCTATGACC TCAAGGGCCC AAGCGTAACA GTGGACACTG 300
CATGCTCCTC GGGTCTCACG GCACTTCATT TAGCATGCCA GAGCATACGC 350
ACTGGTGAGA TCCGACAAGC TTTGGTCGGC GGTGTATACA TTATCCTAAG 400
CCCGGAGAAT ATGATTGCCA TGAGCATGCT GGGGTGATGT CTCCTGTTCC 450
AGAAAGTAAT TGATAAAAGC TAATGCCAGT AGACTGTTTG GCACCGACGG 500
TCTCTCATAC AGCTATGATC ACCGAGCAAC TGGATATGGA CGTGGTGAAG 550
GAGGAGGCAT GATAGTCTTA AAGTCGCTAG ACGACGCGAT GGCAAACGGA 600
GACACAATAC ATGCGGTAAT TCGGCACACA GGGACAAATC AGGATGGTAA 650
GACCAGCGGC CCAACAATGC CCAGTCTGGA AGCCCAGGAG AGACTCATCA 700
AGAAAGTTTA CAGCCAGGCT GGTCTGGATC CATTGGATAC AGAATATGTC 750
GAG 753

(2) INFORMATION FOR SEQ ID NO:74:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Leu Leu Leu Glu Val Thr Tyr Glu Ala Leu Glu Asn Ala Gly Ile
5 10 15

Pro Leu Asn Gln Ile Val Gly Gln Asp Val Gly Val Phe Val Gly
20 25 30

Gly Ser Met Ser Asp Tyr Gln Asn Leu Leu His Lys Asp Ile Ala
35 40 45

Asn Gly Pro Ile Tyr Gln Ala Thr Gly Thr Ala Met Ser Phe Leu
50 55 60

Ala Asn Arg Ile Ser Tyr Ile Tyr Asp Leu Lys Gly Pro Ser Val
65 70 75

Thr Val Asp Thr Ala Cys Ser Ser Gly Leu Thr Ala Leu His Leu
80 85 90

Ala Cys Gln Ser Ile Arg Thr Gly Glu Ile Arg Gln Ala Leu Val
95 100 105

Gly Gly Val Tyr Ile Ile Leu Ser Pro Glu Asn Met Ile Ala Met

```
                110                 115                 120

Ser Met Leu Gly Leu Phe Gly Thr Asp Gly Leu Ser Tyr Ser Tyr
                125                 130                 135

Asp His Arg Ala Thr Gly Tyr Gly Arg Gly Glu Gly Gly Gly Met
                140                 145                 150

Ile Val Leu Lys Ser Leu Asp Asp Ala Met Ala Asn Gly Asp Thr
                155                 160                 165

Ile His Ala Val Ile Arg His Thr Gly Thr Asn Gln Asp Gly Lys
                170                 175                 180

Thr Ser Gly Pro Thr Met Pro Ser Leu Glu Ala Gln Glu Arg Leu
                185                 190                 195

Ile Lys Lys Val Tyr Ser Gln Ala Gly Leu Asp Pro Leu Asp Thr
                200                 205                 210

Glu Tyr Val Glu
```

(2) INFORMATION FOR SEQ ID NO:75:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 692
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

```
AATGCTGCTT GAGGTAGTCT ATGAGGCGTT AGAAGACGGT AAGTCTAACG 50
AATTTCAATC AGTGGTCCTG AGCTAATTGC GATCAAGCTG GCATTACGCT 100
CGACGACATT AAGGGTTCCC AGACATCTGT CTACTGTGGG AGCTTCACCA 150
ACGACTACCG TGAAATGCTG AACAAAGATT TGGGGTACTA CCCCAAGTAC 200
ATGGCCACTG GTGTTGGAAA CTCCATCTTA GCCAACCGCA TTTCATATTT 250
CTATGACCTA CACGGACCAA GTGTGACTGT CGACACAGCC TGCTCTCTTC 300
CCCTGGTCTC ATTCCATATG GGCAACAGAT CAATCCMAGA TGGAGATGCT 350
GACATCTCAA TCGTCATTGG ATCTTCGCTC CATTTTGATC CCAACATGTT 400
CGTCACTATG ACGGACCTTG GGTTTCTCTC AACCGACGGC AGATGCCGTG 450
CTTTTGACGC TAGCGGAAAG GGGTATGTCC GCGGTGAGGG CATCTGCGCT 500
GTTGTTTTGA AACAAAAATC ACGCGCTGAA CTTCACGACA ACAACGTTCG 550
ATCCGTCATT CGTGGCTCGG ATGTCAACCA CGACGGTGCC AAAGACGGTA 600
TCACAATGCC AAACTCGAAG GCTCAGGAGA GCCTCATCAG AAAGACCTAC 650
AAAAACGCTG GACTGAGTAC AAACGACACC CAGTACTTTG AG          692
```

(2) INFORMATION FOR SEQ ID NO:76:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Met Leu Leu Glu Val Val Tyr Glu Ala Leu Glu Asp Ala Gly Ile
5 10 15

Thr Leu Asp Asp Ile Lys Gly Ser Gln Thr Ser Val Tyr Cys Gly
20 25 30

Ser Phe Thr Asn Asp Tyr Arg Glu Met Leu Asn Lys Asp Leu Gly
35 40 45

Tyr Tyr Pro Lys Tyr Met Ala Thr Gly Val Gly Asn Ser Ile Leu
50 55 60

Ala Asn Arg Ile Ser Tyr Phe Tyr Asp Leu His Gly Pro Ser Val
65 70 75

Thr Val Asp Thr Ala Cys Ser Leu Pro Leu Val Ser Phe His Met
80 85 90

Gly Asn Arg Ser Ile Xaa Asp Gly Asp Ala Asp Ile Ser Ile Val
95 100 105

Ile Gly Ser Ser Leu His Phe Asp Pro Asn Met Phe Val Thr Met
110 115 120

Thr Asp Leu Gly Phe Leu Ser Thr Asp Gly Arg Cys Arg Ala Phe
125 130 135

Asp Ala Ser Gly Lys Gly Tyr Val Arg Gly Glu Gly Ile Cys Ala
140 145 150

Val Val Leu Lys Gln Lys Ser Arg Ala Glu Leu His Asp Asn Asn
155 160 165

Val Arg Ser Val Ile Arg Gly Ser Asp Val Asn His Asp Gly Ala
170 175 180

Lys Asp Gly Ile Thr Met Pro Asn Ser Lys Ala Gln Glu Ser Leu
185 190 195

Ile Arg Lys Thr Tyr Lys Asn Ala Gly Leu Ser Thr Asn Asp Thr
200 205 210

Gln Tyr Phe Glu

(2) INFORMATION FOR SEQ ID NO:77:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

TATTTTATTG GAGACAACAT ACGAAGCACT TGAAAATAGT GAGTAAGCCA 50
TGACCGTATT AAGTAAAAGC TCACGAACAG TAAAGGTGGC ACCCCTCTGG 100
CTAGCATTCG CGGCCAAAAT GTAGGCGTTT ACGTTGGTGC ATCCATGTCA 150
GACTACAACG AGCTTTTCGC AAAGGACCCG GATACCAATT TGACATATCG 200
TATTACCGGA ACTGCATCAA ATATTTTGTC AAATCGACTC TCCTACATGT 250
TCGACCTTCA CGGGCCAAGT TTCACGGTGG ACACTGCGTG CTCATCAAGC 300
TTGGCCGCAT TCCATCTGGC CTGTCAGAGT TTGAAGACGG GAGAGGTCCG 350
GCAAGCCATC GTGGGCGGGG CTTACCTTGT ATTATCCCCA GATCCTACGA 400
TCGGAATGAG CAAACTCAGG CTTTACGGCG AACATGGTCG CTCATACACT 450
TACGATCACC GAGGGACTGG ATACGGTCGT GGCGAGGGCG TCGCTAGCCT 500
AATTCTTAAG CCTTTACAAG ATGCTATCGA CGTGGGTGAT ACAATTCGAG 550
CAATCATACG TAACACTGGA ATGAATCAAG ACGGGAAGAC GAACGGAATT 600
ACGCTCCCAA GCAAAGACGC CCAAGAAAGC CTCATAAGGT CTGTCTACAC 650
AGCTGCAGGT CTCGATCCAC TGTATACTTC CTACGTTGAG 690

(2) INFORMATION FOR SEQ ID NO:78:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Ile Leu Leu Glu Thr Thr Tyr Glu Ala Leu Glu Asn Ser Gly Thr
5 10 15

Pro Leu Ala Ser Ile Arg Gly Gln Asn Val Gly Val Tyr Val Gly
20 25 30

Ala Ser Met Ser Asp Tyr Asn Glu Leu Phe Ala Lys Asp Pro Asp
35 40 45

Thr Asn Leu Thr Tyr Arg Ile Thr Gly Thr Ala Ser Asn Ile Leu
50 55 60

Ser Asn Arg Leu Ser Tyr Met Phe Asp Leu His Gly Pro Ser Phe
65 70 75

Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Ala Ala Phe His Leu
80 85 90

Ala Cys Gln Ser Leu Lys Thr Gly Glu Val Arg Gln Ala Ile Val
95 100 105

Gly Gly Ala Tyr Leu Val Leu Ser Pro Asp Pro Thr Ile Gly Met
110 115 120

Ser Lys Leu Arg Leu Tyr Gly Glu His Gly Arg Ser Tyr Thr Tyr

125 130 135

Asp His Arg Gly Thr Gly Tyr Gly Arg Gly Glu Gly Val Ala Ser
140 145 150

Leu Ile Leu Lys Pro Leu Gln Asp Ala Ile Asp Val Gly Asp Thr
155 160 165

Ile Arg Ala Ile Ile Arg Asn Thr Gly Met Asn Gln Asp Gly Lys
170 175 180

Thr Asn Gly Ile Thr Leu Pro Ser Lys Asp Ala Gln Glu Ser Leu
185 190 195

Ile Arg Ser Val Tyr Thr Ala Ala Gly Leu Asp Pro Leu Tyr Thr
200 205 210

Ser Tyr Val Glu

(2) INFORMATION FOR SEQ ID NO:79:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:761
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GCGAATGCTA GAGACGGCTT ATCACGCTCT GGAGGACGGT AAGTCTAACC 50
AGTGCAAATT TAGGGGCTAT AATCTTGGTG TGTGAGAATA ACATACCATC 100
AGCGAGCATC CCCCTGGAGA AGTGCTTCGG CTCAGACACT TCCGTTTATA 150
CCGGGTGCTT CACCAACGAT TATCTCAGCA TACTGCAGCA AGACTTTGAG 200
GCTGAGCAAA GGCACGCAGC CATGGGAATC GCGCCCTCCA TGTTGGCCAA 250
TCGCCTAAGC TGGTTCTTCA ACTTCAAGGG GACATCGATG AACCTGGATT 300
CGGCCTGCTC CAGCAGTCTG GTTGCACTGC ATCTTGCTTC ACAGGACCTC 350
CGTGCTGGTA CCACATCGAT GGTATGTATC GATCATAAAA TCACGTACTC 400
CTTCATTAAT AAATAAATGT TTTAGGCACT AGTTGGAGGG GCGAATCTTG 450
TCTACCACCC CGACTTCATG GAGATGATGT CAAACTTCAA CTTCCTGTCT 500
CCCGACAGCC GTTCTTGGAG TTTCGATCAA CGTGCTAATG GTTATGCGCG 550
TGGGGAAGGA ACCGCCGTGA TGGTCGTCAA ACGCCTTGCA GATGCACTGC 600
GAGATGGAGA TACAATCAGA ACCGTAATCT GGAGTACCGG GTCGAACCAA 650
GACGGGAGAA CACCTGGGAT CACGCAGCCA AGTAAAGAAG CGCAGTTAAA 700
TCTCATCGAG CGCACCTACA AACAAGCGAA GATTGATATG GAGCCTACCA 750
GATTCTTCGA G 761

(2) INFORMATION FOR SEQ ID NO:80:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 214
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein

(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Arg Met Leu Glu Thr Ala Tyr His Ala Leu Glu Asp Ala Ser Ile
5 10 15

Pro Leu Glu Lys Cys Phe Gly Ser Asp Thr Ser Val Tyr Thr Gly
20 25 30

Cys Phe Thr Asn Asp Tyr Leu Ser Ile Leu Gln Gln Asp Phe Glu
35 40 45

Ala Glu Gln Arg His Ala Ala Met Gly Ile Ala Pro Ser Met Leu
50 55 60

Ala Asn Arg Leu Ser Trp Phe Phe Asn Phe Lys Gly Thr Ser Met
65 70 75

Asn Leu Asp Ser Ala Cys Ser Ser Ser Leu Val Ala Leu His Leu
80 85 90

Ala Ser Gln Asp Leu Arg Ala Gly Thr Thr Ser Met Ala Leu Val
95 100 105

Gly Gly Ala Asn Leu Val Tyr His Pro Asp Phe Met Glu Met Met
110 115 120

Ser Asn Phe Asn Phe Leu Ser Pro Asp Ser Arg Ser Trp Ser Phe
125 130 135

Asp Gln Arg Ala Asn Gly Tyr Ala Arg Gly Glu Gly Thr Ala Val
140 145 150

Met Val Val Lys Arg Leu Ala Asp Ala Leu Arg Asp Gly Asp Thr
155 160 165

Ile Arg Thr Val Ile Trp Ser Thr Gly Ser Asn Gln Asp Gly Arg
170 175 180

Thr Pro Gly Ile Thr Gln Pro Ser Lys Glu Ala Gln Leu Asn Leu
185 190 195

Ile Glu Arg Thr Tyr Lys Gln Ala Lys Ile Asp Met Glu Pro Thr
200 205 210

Arg Phe Phe Glu

(2) INFORMATION FOR SEQ ID NO:81:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1221
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

```
AAGGAGGGGC CGCCCGGGAG AAGAAGTTAT CGTGGGCGCC GATTCGGTCG 50
ACCGGCAGCA ATTGCAGCCA GATTGCCGCG AGGGCTTCCT CCATTCCCGG 100
CGCGGGCGCA ACGAATCCGG TGTACTCCAG ATGCCGTGCG GTCCGGGGGA 150
GAGCTGCCTG ATCCAGTTTG AGATTCTTGT TTAAAGGAAG TTCGGCCAGC 200
TTCTCTATGG CGGCGGGGAC CATGTGAGCG GGGAGCAGAG CCTTCATGTG 250
CTGGCGAATC GTTTCCGTGG ACGCTCCGCC GACTGCATAC GCCGCGAGAT 300
ACTTCTCGCC GGGGATATCG TCTCGGACCA GCACAACGCC GTCCGTGACG 350
CCCGGGCACG ACTGCAGCGC GGCCTGAATT TCGCCGAGTT CTATGCGATG 400
CCCGCGAAGC TTGATCTGGC CGTCGTTTCT GCCCAGAAAA TCGATGCGCC 450
CATCCGGCAG ATAGCGCGCG CGATCGCCCG TGCGGTACAT ACGCGCGCCC 500
GGAAATGGGC TAAACGGGTT CGGCACAAAG TAGGCTGCGG TGAGATCGCT 550
GCGCCCCGCA TAGCCGCGCG CGACACCGTC TCCGGCAGCG TACAGCCAGC 600
CTTCCACTCC CGGCGGAACG GGAGCGAATT GCTCGTCGAG CACGTAGGTT 650
TGGACGTTCG AAATTGGACG GCCGATGGGA ATCGACGGGG TCCCGGCGGG 700
GACCGAATCG ATGACGCCAC ACGCCGTGAG CATCGTGTTC TCGGTAGGGC 750
CGTAACCGTT CAAGAGGCGG GCGGGCTTGC CGTGCTCGAT CACCATGCGC 800
ATCCAGTGGG GATCCAGCGC TTCGCCGCCG ACAATCACAT TGGTCAGCGA 850
TTCGAATCCG GCTGGATCTT CGCGGGCAAC CTGATTGAAC AGAGATGCAG 900
TAAGGATAAT CGTGTCCACG TGGAAGCGGC GAAAGGCGAG AATCAGCTCG 1000
CGGGGCGCCA TCAAGGTCTC TTTCGAAAGA ACGACGATTC GCGCGCCATG 1050
CAGCAGGCCG CCCCATAACT CGAAGGTGGG AGGGTCGAAA CCGAAGGCCG 1100
ACATCTGTCC CACGGTATCG GCGGGTGAGA ATTGTACGTA GTTGGTCCGG 1150
CTAACGAGGT TGACAATCGC CCCGTGGGGG ACGGCGACCC CCTTGGGCTT 1200
GCCGGTCGTG CCGGACGTGT A                                  1221
```

(2) INFORMATION FOR SEQ ID NO:82:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

```
Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Val Ala Val Pro
 5                  10                  15

His Gly Ala Ile Val Asn Leu Val Ser Arg Thr Asn Tyr Val Gln
            20                  25                  30

Phe Ser Pro Ala Asp Thr Val Gly Gln Met Ser Ala Phe Gly Phe
            35                  40                  45

Asp Pro Pro Thr Phe Glu Leu Trp Gly Gly Leu Leu His Gly Ala
            50                  55                  60

Arg Ile Val Val Leu Ser Lys Glu Thr Leu Met Ala Pro Arg Glu
            65                  70                  75

Leu Ile Leu Ala Phe Arg Arg Phe His Val Asp Thr Ile Ile Leu
            80                  85                  90
```

Thr Ala Ser Leu Phe Asn Gln Val Ala Arg Glu Asp Pro Ala Gly
95 100 105

Phe Glu Ser Leu Thr Asn Val Ile Val Gly Gly Glu Ala Leu Asp
110 115 120

Pro His Trp Met Arg Met Val Ile Glu His Gly Lys Pro Ala Arg
125 130 135

Leu Leu Asn Gly Tyr Gly Pro Thr Glu Asn Thr Met Leu Thr Ala
140 145 150

Cys Gly Val Ile Asp Ser Val Pro Ala Gly Thr Pro Ser Ile Pro
155 160 165

Ile Gly Arg Pro Ile Ser Asn Val Gln Thr Tyr Val Leu Asp Glu
170 175 180

Gln Phe Ala Pro Val Pro Pro Gly Val Glu Gly Trp Leu Tyr Ala
185 190 195

Ala Gly Asp Gly Val Ala Arg Gly Tyr Ala Gly Arg Ser Asp Leu
200 205 210

Thr Ala Ala Tyr Phe Val Pro Asn Pro Phe Ser Pro Phe Pro Gly
215 220 225

Ala Arg Met Tyr Arg Thr Gly Asp Arg Ala Arg Tyr Leu Pro Asp
230 235 240

Gly Arg Ile Asp Phe Leu Gly Arg Asn Asp Gly Gln Ile Lys Leu
245 250 255

Arg Gly His Arg Ile Glu Leu Gly Glu Ile Gln Ala Ala Leu Gln
260 265 270

Ser Cys Pro Gly Val Thr Asp Gly Val Val Leu Val Arg Asp Asp
275 288 285

Ile Pro Gly Glu Lys Tyr Leu Ala Ala Tyr Ala Val Gly Gly Ala
290 295 300

Ser Thr Glu Thr Ile Arg Gln His Met Lys Ala Leu Leu Pro Ala
305 310 315

His Met Val Pro Ala Ala Ile Glu Lys Leu Ala Glu Leu Pro Leu
320 325 330

Asn Lys Asn Leu Lys Leu Asp Gln Ala Ala Leu Pro Arg Thr Ala
335 340 345

Arg His Leu Glu Tyr Thr Gly Phe Val Ala Pro Ala Pro Gly Met
350 355 360

Glu Glu Ala Leu Ala Ala Ile Trp Leu Gln Leu Leu Pro Val Asp

365 370 375

Arg Ile Gly Ala His Asp Asn Phe Phe Ser Arg Ala Ala Pro Pro
380 385 390

(2) INFORMATION FOR SEQ ID NO:83
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:1222
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CGTTTCACCC CAAGAATCTC AGACCATATA TCAGCAATGG CCTTCTCCCT 50
GGCATTGCCC GGAGCGACAT AGATCGGATC CCGAATCACA GTATCGCGAT 100
CAAATGGCGG CAGGGCGTTT CGGTCAATCT TGCCGTTCGG CGTTAAAGGG 150
AGAGAATCGA CAATGACGAA GGCGCTGGGC ACCATGTAGT CCGGCAGTTT 200
TGCCTTCAGA TGGGCGCGCA ATTCGCTTAT TTCGGGAGCA CCTTCCCGTG 250
CGACGATATA AGCAACTAAT TGCTTTTCTT CGCTAGGGTC TTTTGTCGTT 300
GTGACCACAG CTTCTCGAAT CGGGGATGTT GCGCAACAGG ACTTCGATTT 350
CTCCAGCTCG ATGCGATAGC CGCGAATCTT GACCTGATTG TCGGTGCGGC 400
CGATAAACTC GATGTTGCCA TCCGGCAAAT AACGCGCAAG ATCGCCAGTT 450
CGATAGAGGC GCTGCGCTGG CTCGCGATCG AATGAATGGT AGATGAACCT 500
CTCCGCCGTC AGTTCCGGCC GGTTGAGATA CCCTCGCGCC AGTCCGTCGC 550
CGCCAATGTA GATCTCTCCA ACCACGCCGA TCGGCACCGG ATTGAGATGA 600
GCATCCAGTA TGTAGATCTG CGTATTCGCG ATCGGTCGGC CAATGGGCGG 650
TAATTCTCCC CAGCACTCTG GCGGACCGTC CACAGTAAAC GCTGTCACAA 700
CGTGGCTTTC CGTCGGCCCA TACTGGTTGA CCAAATGACA CTCGGGCAAC 750
GTGTCAAGGA AACTTCTGAT CCGCGGCGTT ATCTGCAGCC GCTCTCCCGC 800
CGTAATGACT TCGCGCAGCT GCGGCAAAAC CACATTCTCC ATGTGCGCGG 850
CTTCCGCCAT CTGTTGCAGT ACGACAAAAG GCACAAAAAG TCTCTCTACT 900
CGCTTCATTC GCAGGAAATT CAACAGGGCT GGCGGATCGC GTCGGATTTG 950
CGCGGGCAGT AGCACCAGTG TGCCTCCTGA GCACCACGTG CTAAACATCT 1000
CTTGAAACGA AACATCGAAA CTCAACGAGG CAAACTGTAA CGTTCGCGCC 1050
GGCACCGAAC GAGAAAAATC CTCAATTTGC CACGCGATCA GGTTGGCAAG 1100
CGCGCGGTGT TCCATCACCA CACCCTTCGG CTTGCCCGTC GTGCCAATCC 1150
CGCGGCCATG GCGGCCGGGA GCATGCGACG TCGGGCCCAA TTCGCCCTAT 1200
AGTGAGTCGT ATTACAATTC AA 1222

(2) INFORMATION FOR SEQ ID NO:84
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Gly Thr Thr Gly Lys Pro Lys Gly Val Val Met Glu His Arg Ala
5 10 15

Leu Ala Asn Leu Ile Ala Trp Gln Ile Glu Asp Phe Ser Arg Ser
20 25 30

Val Pro Ala Arg Thr Leu Gln Phe Ala Ser Leu Ser Phe Asp Val
35 40 45

Ser Phe Gln Glu Met Phe Ser Thr Trp Cys Ser Gly Gly Thr Leu
50 55 60

Val Leu Leu Pro Ala Gln Ile Arg Arg Asp Pro Pro Ala Leu Leu
65 70 75

Asn Phe Leu Arg Met Lys Arg Val Glu Arg Leu Phe Val Pro Phe
80 85 90

Val Val Leu Gln Gln Met Ala Glu Ala Ala His Met Glu Asn Val
95 100 105

Val Leu Pro Gln Leu Arg Glu Val Ile Thr Ala Gly Glu Arg Leu
110 115 120

Gln Ile Thr Pro Arg Ile Arg Ser Phe Leu Asp Thr Leu Pro Glu
125 130 135

Cys His Leu Val Asn Gln Tyr Gly Pro Thr Glu Ser His Val Val
140 145 150

Thr Ala Phe Thr Val Asp Gly Pro Pro Glu Cys Trp Gly Glu Leu
155 160 165

Pro Pro Ile Gly Arg Pro Ile Ala Asn Thr Gln Ile Tyr Ile Leu
170 175 180

Asp Ala His Leu Asn Pro Val Pro Ile Gly Val Val Gly Glu Ile
185 190 195

Tyr Ile Gly Gly Asp Gly Leu Ala Arg Gly Tyr Leu Asn Arg Pro
200 205 210

Glu Leu Thr Ala Glu Arg Phe Ile Tyr His Ser Phe Asp Arg Glu
215 220 225

Pro Ala Gln Arg Leu Tyr Arg Thr Gly Asp Leu Ala Arg Tyr Leu
230 235 240

Pro Asp Gly Asn Ile Glu Phe Ile Gly Arg Thr Asp Asn Gln Val
245 250 255

Lys Ile Arg Gly Tyr Arg Ile Glu Leu Glu Lys Ser Lys Ser Cys
260 265 270

Cys Ala Thr Ser Pro Ile Arg Glu Ala Val Val Thr Thr Thr Lys

275 288 285

Asp Pro Ser Glu Glu Lys Gln Leu Val Ala Tyr Ile Val Ala Arg
290 295 300

Glu Gly Ala Pro Glu Ile Ser Glu Leu Arg Ala His Leu Lys Ala
305 310 315

Lys Leu Pro Asp Tyr Met Val Pro Ser Ala Phe Val Ile Val Asp
320 325 330

Ser Leu Pro Leu Thr Pro Asn Gly Lys Ile Asp Arg Asn Ala Leu
335 340 345

Pro Pro Phe Asp Arg Asp Thr Val Ile Arg Asp Pro Ile Tyr Val
350 355 360

Ala Pro Gly Asn Ala Arg Glu Lys Ala Ile Ala Asp Ile Trp Ser
365 370 375

Glu Ile Leu Gly Val Lys Arg Ile Gly Val His Asp Asn Phe Phe
380 385 390

Ala Pro Gly Gly Pro Ser
395

(2) INFORMATION FOR SEQ ID NO:85
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:1200
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

AATCTACACG TCCGGCACCA CCGGCAAGCC CAAGGGGGCC ATAATCCATC 50
ACCTGGGACT GGCGAATTAC TTGGTGTGGT GCTCGCGGGC TTACGCGATT 100
GCTCAAGGAG TGGGAGCACC GGTCCACTCG TCGATCTCGT TCGATCTGAC 150
GATCACTGCC TTGCTTGCCC CCTTGGTCGT CGGCCGGCGC ATCGACCTGC 200
TTGATGAAGA ACTGGGCATC GAGCAACTGA GTTACGCTCT CCGGCGATCG 250
CGCGACTATA GCCTGGTCAA GATCACTCCG GCTCACCTGC GCTGGCTCGG 300
CGATGAACTG GGACCCTGCG AGGCCGAAGG TCGTACGCGA GCTTTCATCA 350
TCGGTGGTGA GCAACTGACG GCCGAACACG TCKCATTCTG GAGGCGGCAC 400
GCGCCGGGGA CGAGCCTGAT CAACGAGTAT GGTCCGACCG AGACGGTCGT 450
CGGCTGCTGC GTGTACCGCG TGCCTCCTGA CCAGGAGATT TCGGGGCCCA 500
TCCCGATTGG CCGACCGATC GCCAACACGC GTCTCTACGT CCTCGATCCG 550
GATCTCGCGC TGGTACCCAT CGGCGTTGCA GGCGAGCTGT ACATCGGCGG 600
TGCCGGGGTC GCGCGGGGGT ATCTCAACAG GCCCGGCCTG ACCGCTGAAA 650
GGTTCATCCC CGACCCGTTC GGCAAGAAGC CGGGCGAGCG CCTCTATCGC 700
ACCGGAGACC TCGCCCGATG GCGGTCCGAC GGTAACCTCG AGTATCTCGG 750
CAGGGTCGAT CGCCAGGTTA AAGTCCGCGG GTTTCGGATC GAACCCGGGG 800
AGATCGAACA GGCACTCGCC CGGCACTCCG CGGTACGCGA GTCCGTCGTG 850
GTCGCAAGCG CAGGTGCATC GGACGTGCAA CGCCTCGTCG CCTATCTGGT 900

TCTTGCGGAG GCAGGGCCGG CACCGCCCGA CTCGGAGCTG CGCGAGTTCC 950
TGCGGACGTT ACTCCCCGAG CCGATGATAC CCTCGGCATT CGTTGTGCTG 1000
GAGACGCTCC CACTGACCCA CAACGGGAAG GTGGACCGAG AGGCCCTGCC 1050
GGCCCCTGAG GGTGTGCCCT TCCGTGGGGA TGCTCGTTTC GTTGCTCCCC 1100
GCGGCCCGCT CGAACAGGAG GTGGCATCGA TCTGGGGTGC AGTCCTCGGA 1150
CTGGAGCGTA TCGGCGCCCT TGACAACTTC TTCTTCCCTC GGCGGCCCCT 1200

(2) INFORMATION FOR SEQ ID NO:86
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Ala Ile Ile
5 10 15

His His Leu Gly Leu Ala Asn Tyr Leu Val Trp Cys Ser Arg Ala
20 25 30

Tyr Ala Ile Ala Gln Gly Val Gly Ala Pro Val His Ser Ser Ile
35 40 45

Ser Phe Asp Leu Thr Ile Thr Ala Leu Leu Ala Pro Leu Val Val
50 55 60

Gly Arg Arg Ile Asp Leu Leu Asp Glu Glu Leu Gly Ile Glu Gln
65 70 75

Leu Ser Tyr Ala Leu Arg Arg Ser Arg Asp Tyr Ser Leu Val Lys
80 85 90

Ile Thr Pro Ala His Leu Arg Trp Leu Gly Asp Glu Leu Gly Pro
95 100 105

Cys Glu Ala Glu Gly Arg Thr Arg Ala Phe Ile Ile Gly Gly Glu
110 115 120

Gln Leu Thr Ala Glu His Val Xaa Phe Trp Arg Arg His Ala Pro
125 130 135

Gly Thr Ser Leu Ile Asn Glu Tyr Gly Pro Thr Glu Thr Val Val
140 145 150

Gly Cys Cys Val Tyr Arg Val Pro Pro Asp Gln Glu Ile Ser Gly
155 160 165

Pro Ile Pro Ile Gly Arg Pro Ile Ala Asn Thr Arg Leu Tyr Val
170 175 180

Leu Asp Pro Asp Leu Ala Leu Val Pro Ile Gly Val Ala Gly Glu
185 190 195

TCTTGCGGAG GCAGGGCCGG CACCGCCCGA CTCGGAGCTG CGCGAGTTCC 950
TGCGGACGTT ACTCCCCGAG CCGATGATAC CCTCGGCATT CGTTGTGCTG 1000
GAGACGCTCC CACTGACCCA CAACGGGAAG GTGGACCGAG AGGCCCTGCC 1050
GGCCCCTGAG GGTGTGCCCT TCCGTGGGGA TGCTCGTTTC GTTGCTCCCC 1100
GCGGCCCGCT CGAACAGGAG GTGGCATCGA TCTGGGGTGC AGTCCTCGGA 1150
CTGGAGCGTA TCGGCGCCCT TGACAACTTC TTCTTCCCTC GGCGGCCCCT 1200

(2) INFORMATION FOR SEQ ID NO:86
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Ala Ile Ile
5 10 15

His His Leu Gly Leu Ala Asn Tyr Leu Val Trp Cys Ser Arg Ala
20 25 30

Tyr Ala Ile Ala Gln Gly Val Gly Ala Pro Val His Ser Ser Ile
35 40 45

Ser Phe Asp Leu Thr Ile Thr Ala Leu Leu Ala Pro Leu Val Val
50 55 60

Gly Arg Arg Ile Asp Leu Leu Asp Glu Glu Leu Gly Ile Glu Gln
65 70 75

Leu Ser Tyr Ala Leu Arg Arg Ser Arg Asp Tyr Ser Leu Val Lys
80 85 90

Ile Thr Pro Ala His Leu Arg Trp Leu Gly Asp Glu Leu Gly Pro
95 100 105

Cys Glu Ala Glu Gly Arg Thr Arg Ala Phe Ile Ile Gly Gly Glu
110 115 120

Gln Leu Thr Ala Glu His Val Xaa Phe Trp Arg Arg His Ala Pro
125 130 135

Gly Thr Ser Leu Ile Asn Glu Tyr Gly Pro Thr Glu Thr Val Val
140 145 150

Gly Cys Cys Val Tyr Arg Val Pro Pro Asp Gln Glu Ile Ser Gly
155 160 165

Pro Ile Pro Ile Gly Arg Pro Ile Ala Asn Thr Arg Leu Tyr Val
170 175 180

Leu Asp Pro Asp Leu Ala Leu Val Pro Ile Gly Val Ala Gly Glu
185 190 195

Leu Tyr Ile Gly Gly Ala Gly Val Ala Arg Gly Tyr Leu Asn Arg
200 205 210

Pro Gly Leu Thr Ala Glu Arg Phe Ile Pro Asp Pro Phe Gly Lys
215 220 225

Lys Pro Gly Glu Arg Leu Tyr Arg Thr Gly Asp Leu Ala Arg Trp
230 235 240

Arg Ser Asp Gly Asn Leu Glu Tyr Leu Gly Arg Val Asp Arg Gln
245 250 255

Val Lys Val Arg Gly Phe Arg Ile Glu Pro Gly Glu Ile Glu Gln
260 265 270

Ala Leu Ala Arg His Ser Ala Val Arg Glu Ser Val Val Val Ala
275 288 285

Ser Ala Gly Ala Ser Asp Val Gln Arg Leu Val Ala Tyr Leu Val
290 295 300

Leu Ala Glu Ala Gly Pro Ala Pro Pro Asp Ser Glu Leu Arg Glu
305 310 315

Phe Leu Arg Thr Leu Leu Pro Glu Pro Met Ile Pro Ser Ala Phe
320 325 330

Val Val Leu Glu Thr Leu Pro Leu Thr His Asn Gly Lys Val Asp
335 340 345

Arg Glu Ala Leu Pro Ala Pro Glu Gly Val Pro Phe Arg Gly Asp
350 355 360

Ala Arg Phe Val Ala Pro Arg Gly Pro Leu Glu Gln Glu Val Ala
365 370 375

Ser Ile Trp Gly Ala Val Leu Gly Leu Glu Arg Ile Gly Ala Leu
380 385 390

Asp Asn Phe Phe Phe Pro Arg Arg Pro
395

(2) INFORMATION FOR SEQ ID NO:87:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1204
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

AGGGGCCGCC GGGCGAGAAG AAGTTCGCGG TGATGCTCAC CGGCGCGTCG 50
AGCTTCAACG CCTCCTGCCA GATCTCCGCG AGCTTGCTCT CCGTCTCCGT 100

GCCCGGCGCT ACGTATTGGG CGCCGGCGCT ACGGTCGATG GACGGCAGCG 150
CCTTACGATC GATCTTGCCG TTGGCATTCA GCGGAAAGGC CTCCAGGACG 200
CGCCAGCCGC TGGGAATCAT GTACTCGGGC AGGGCCAGCT TGAGGCGCAT 250
CCGCAGCGCC GAGATGAGCA CCTCTTCGTC CGCGGTCTGG GCCACGACGT 300
AGGCGACGAG GGCCTTGTTC TCCCCCTCTC CCTGCGCCAC GACCAGGGCG 350
TCGTCGACGC CAGCCTCGGT CTTCAGCGCG GTCTCGATCT CGCCGAGCTC 400
GATGCGGAAG CCGCGGATCT TGATCTGGTC GTCGAGGCGG CCGAGGAACT 450
CGAGATCGCC GCTGGCGAGC CGGCGAACGA GGTCGCCGCT GCGATAGAGG 500
CGCCCTTCGC CGAAGGGATT GGCGATGAAC TTCGCCGCCG TCAGCTCCGG 550
CTGGTTGACG TAGCCTCTGG CCACCCCTGC CCCGCCAATG CACAGCTCGC 600
CGGCCACGCC GACCGGCGCG ATCTCCAGTG CCTCGTTGAG GACATACAGC 650
TCCGTGTTGT CCATGGCCCT GCCGATGGGC AGGCGCTCCG GCAGGCCGGC 700
CTGGAGAGCG GCGGTGACGT CGAACATGGC GCAGCCGACC ACGGTCTCCG 750
TGGGACCGTA GTGGTTGTAG ATCTGGGCGT GGGGGAAGCG CGTTTGCAGC 800
TCGCGGGCGA GCGAGGCGGG AAACGATTCG CCGCCGATGA CGAAAACGTG 850
TTGAGATGAA GCCCGGGCCG TGTCTTCCGT CAGCTCCGCG CTGTCGAGCA 900
GAGCGAGCAT ACCGGTGAGA TGCATCGGCG TCATGCGCAG CAGATAAGCC 950
CGTTCGTCGC CGGCCAACGC TTTCGCGAGC TCGTTCAACT CATCGCCGGG 1000
CGTGGTCAGC GAGACGCAGC CACCCCGGAG CAAGGGAACA TACAGGCTGG 1050
GCACGGTGAT GTCGAAGCCG TGGGAGGTGA CGACGAGGGA GCCGGCCAAC 1100
CCCTTCGCGT AGTAGCGCTG CGAAGCGAAG GCGCAGTAGT CACTGAGGCC 1150
GGCGTGTCTG ATCTCCACGC CCTTCGGCTT GCCCGTCGTG CCGGACGTGT 1200
AGAT 1204

(2) INFORMATION FOR SEQ ID NO:88:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 401
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Val Glu Ile
5 10 15

Arg His Ala Gly Leu Ser Asp Tyr Cys Ala Phe Ala Ser Gln Arg
20 25 30

Tyr Tyr Ala Lys Gly Leu Ala Gly Ser Leu Val Val Thr Ser His
35 40 45

Gly Phe Asp Ile Thr Val Pro Ser Leu Tyr Val Pro Leu Leu Arg
50 55 60

Gly Gly Cys Val Ser Leu Thr Thr Pro Gly Asp Glu Leu Asn Glu
65 70 75

Leu Ala Lys Ala Leu Ala Gly Asp Glu Arg Ala Tyr Leu Leu Arg
80 85 90

Met Thr Pro Met His Leu Thr Gly Met Leu Ala Leu Leu Asp Ser
95 100 105

Ala Glu Leu Thr Glu Asp Thr Ala Arg Ala Ser Ser Gln His Val
110 115 120

Phe Val Ile Gly Gly Glu Ser Phe Pro Ala Ser Leu Ala Arg Glu
125 130 135

Leu Gln Thr Arg Phe Pro His Ala Gln Ile Tyr Asn His Tyr Gly
140 145 150

Pro Thr Glu Thr Val Val Gly Cys Ala Met Phe Asp Val Thr Ala
155 160 165

Ala Leu Gln Ala Gly Leu Pro Glu Arg Leu Pro Ile Gly Arg Ala
170 175 180

Met Asp Asn Thr Glu Leu Tyr Val Leu Asn Glu Ala Leu Glu Ile
185 190 195

Ala Pro Val Gly Val Ala Gly Glu Leu Cys Ile Gly Gly Ala Gly
200 205 210

Val Ala Arg Gly Tyr Val Asn Gln Pro Glu Leu Thr Ala Ala Lys
215 220 225

Phe Ile Ala Asn Pro Phe Gly Glu Gly Arg Leu Tyr Arg Ser Gly
230 235 240

Asp Leu Val Arg Arg Leu Ala Ser Gly Asp Leu Glu Phe Leu Gly
245 250 255

Arg Leu Asp Asp Gln Ile Lys Ile Arg Gly Phe Arg Ile Glu Leu
260 265 270

Gly Glu Ile Glu Thr Ala Leu Lys Thr Glu Ala Gly Val Asp Asp
275 288 285

Ala Leu Val Val Ala Gln Gly Glu Gly Glu Asn Lys Ala Leu Val
290 295 300

Ala Tyr Val Val Ala Gln Thr Ala Asp Glu Glu Val Leu Ile Ser
305 310 315

Ala Leu Arg Met Arg Leu Lys Leu Ala Leu Pro Glu Tyr Met Ile
320 325 330

Pro Ser Gly Trp Arg Val Leu Glu Ala Phe Pro Leu Asn Ala Asn
335 340 345

Gly Lys Ile Asp Arg Lys Ala Leu Pro Ser Ile Asp Arg Ser Ala
350 355 360

Gly Ala Gln Tyr Val Ala Pro Gly Thr Glu Thr Glu Ser Lys Leu
365 370 375

Ala Glu Ile Trp Gln Glu Ala Leu Lys Leu Asp Ala Pro Val Ser

380 385 390

Ile Thr Ala Asn Phe Phe Ser Pro Gly Gly Pro
395 400

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:1190
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ATCTACACCT CGGGCACGAC CGGCAAGCCG AAGGGGATCA TGTATTCGCA 50
TCGATACCTG TTGCATAATA TGCGCAACTA CGGCGACTTA TTTCAGGTCT 100
CCCCCCACGA TCGCTGGAGT TGGTTGCATT CCTACAGCTA TGCTTCGGCG 150
AATACTGATA TCCTTTGCCC GCTACTGCAC GGCGCCGCCG TCTGCCCTTG 200
GAATTTGCAT CGTAATGGCC TATCGGGCTT AGCTCGTTGG CTCGCCGAGT 250
CGCGAATCAC CATTTTGAAC TGGATGCCGA CACCGCTACG CAGTTTGGCA 300
AAGCTCTGGC CGCCAAAGCA CGTGCTTCCC GATCTGCGAC TTACAGTGTT 350
GGGCGGCGAA ACGCTGTTTG CCCAAGACGT TGCTGACTTT CGGCGAATAA 400
TTTCGCTGAA TTGCCTAATC GCCAATCGTC TGGGAACTTC GGAAACTGGA 450
TTGTTTCGGC TCGCGTTTCT CGACCGAGAG ACTCCCCTTG CTAATGGTTC 500
CATACAGGCC GGATACGAAG TTCCAGACAA GACCGTCGTC CTGTTCGACG 550
AATATGGAGT TGAGCTGGCC CCTGGCAACG TCGGTCAGAT TGGCGTGCGC 600
AGCAGGTACT TGCCGCCTGG ATACTGGCGA CGGCCGGAGT TGACAAGCGA 650
GCGATTTCTA ACCAGTAAAG GCGATGATGA CGTACGGACC TTCCTCACCG 700
GCGACCTTGG GCGAATGCGG GACGACGGAT GCCTCGAGCA CTGCGGACGG 750
CTCGACTCCC AAGTGAAGAT CCGTGGTCAC CGCATCGCAA TGGGAGAGAT 800
CGAATTCTTG CTTCGGACAT GCGACGGAGT CAGCGAAGCA GTTGTCATTG 850
CCAGGCCACA TTCAGACGGT GAAACCCGTT TGATAGCTTA TTTTGTGCCG 900
ACCGAGAAAA GCGCTATCGA TGTATCGAGC CTTCGTCGGC ACCTGCTGGG 950
AAAGCTGCCT GGCCACATGA TCCCCTCGGC GTTTGTGCGG CTCGACGGCG 1000
TGCCCAAAAA CGCCAACCAA AAAGTAGATT GGGCGGCCTT GCCAGCACCG 1050
AACTTCCAAA ACCAGGGACA GCAGCACGTA CCGCCACAAA CGCCTTGGCA 1100
GCGACATCTC GTGGAGTTGT GGCAAAAGTT GTTGAATGTG GAATCGATCG 1150
GCATCCACGA TGACTTCTTC GCCCTCGGCG GCCCCTCCTT 1190

(2) INFORMATION FOR SEQ ID NO:90:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Ile Met Tyr
5 10 15

Ser His Arg Tyr Leu Leu His Asn Met Arg Asn Tyr Gly Asp Leu

20 25 30

Phe Gln Val Ser Pro His Asp Arg Trp Ser Trp Leu His Ser Tyr
35 40 45

Ser Tyr Ala Ser Ala Asn Thr Asp Ile Leu Cys Pro Leu Leu His
50 55 60

Gly Ala Ala Val Cys Pro Trp Asn Leu His Arg Asn Gly Leu Ser
65 70 75

Gly Leu Ala Arg Trp Leu Ala Glu Ser Arg Ile Thr Ile Leu Asn
80 85 90

Trp Met Pro Thr Pro Leu Arg Ser Leu Ala Lys Leu Trp Pro Pro
95 100 105

Lys His Val Leu Pro Asp Leu Arg Leu Thr Val Leu Gly Gly Glu
110 115 120

Thr Leu Phe Ala Gln Asp Val Ala Asp Phe Arg Arg Ile Ile Ser
125 130 135

Leu Asn Cys Leu Ile Ala Asn Arg Leu Gly Thr Ser Glu Thr Gly
140 145 150

Leu Phe Arg Leu Ala Phe Leu Asp Arg Glu Thr Pro Leu Ala Asn
155 160 165

Gly Ser Ile Gln Ala Gly Tyr Glu Val Pro Asp Lys Thr Val Val
170 175 180

Leu Phe Asp Glu Tyr Gly Val Glu Leu Ala Pro Gly Asn Val Gly
185 190 195

Gln Ile Gly Val Arg Ser Arg Tyr Leu Pro Pro Gly Tyr Trp Arg
200 205 210

Arg Pro Glu Leu Thr Ser Glu Arg Phe Leu Thr Ser Lys Gly Asp
215 220 225

Asp Asp Val Arg Thr Phe Leu Thr Gly Asp Leu Gly Arg Met Arg
230 235 240

Asp Asp Gly Cys Leu Glu His Cys Gly Arg Leu Asp Ser Gln Val
245 250 255

Lys Ile Arg Gly His Arg Ile Ala Met Gly Glu Ile Glu Phe Leu
260 265 270

Leu Arg Thr Cys Asp Gly Val Ser Glu Ala Val Val Ile Ala Arg
275 288 285

Pro His Ser Asp Gly Glu Thr Arg Leu Ile Ala Tyr Phe Val Pro
290 295 300

Thr Glu Lys Ser Ala Ile Asp Val Ser Ser Leu Arg Arg His Leu
305 310 315

Leu Gly Lys Leu Pro Gly His Met Ile Pro Ser Ala Phe Val Arg
320 325 330

Leu Asp Gly Val Pro Lys Asn Ala Asn Gln Lys Val Asp Trp Ala
335 340 345

Ala Leu Pro Ala Pro Asn Phe Gln Asn Gln Gly Gln Gln His Val
350 355 360

Pro Pro Gln Thr Pro Trp Gln Arg His Leu Val Glu Leu Trp Gln
365 370 375

Lys Leu Leu Asn Val Glu Ser Ile Gly Ile His Asp Asp Phe Phe
380 385 390

Ala Leu Gly Gly Pro Ser
395

(2) INFORMATION FOR SEQ ID NO:91:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1178
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

AAGGAGGGGC CGCCCGGCGC GAAGAAGTTC TCGTGTAGCC CGACGCGTTC 50
CAGCTGCAGC ACGGCGCACC AGATCGCTGC GACCTGCCGC TGGACGTCCG 100
TCATGATCGC GGTGTCCGCT GCGGCCGCTG CCGCGCGATT CACCTGTGGA 150
ATGGGCAGGG CCTTGCGGTC GATCTTGTCG TTCGGCGTGA GCGGCAGCGC 200
GGCGAGCGAT ACGATCACCT GTGGCACCAT GTACTCGGGG AGTCTCGCGC 250
GGAGCGCCGT CCGGAGCTCG TCGAGCGGCA GCACGCCGTC TTCTGCCGGG 300
ACGACGTACG CCACCAGACG CTGATCGCCG GGGGTGTCCT CGCGCACGAC 350
GGCCACGCTG CGGCGCACCG ACGGATGCTC GGACAGGACC GATTCGATCT 400
CCCCCAGCTC GATCCGGTAG CCGCGAAGCT TCACCTGATG ATCTCGGCGT 450
CCGACGAACT CGAGGGCCCG ATCGGCGCGC AGTCGTACGA TGTCGCCGGT 500
GCGGTACACG CGCTCCGCCG GTCTGCCCGC GACCTCGACG ACGACGAACT 550
TTTCTGCCGT GAGCTCGGGT CGATGACGAT AGCCCCGCGC CACGCCCTCT 600
CCTCCGATGC ACAGCTCACC CGGCACGCCG ATGGGAGCCT GGCGACCCGC 650
GGCGTCGAGC ACGTAGACGT TCGTGTTGGC GATGGGATGG CCGATCGGAA 700
TATCGCGATC GCAATCCGTG ACCTGATGCA CGGTCGACCA GATCGTCGTC 750
TCGGTCGGGC CGTACATGTT CCACAGCGCC CGCACCCTCG ACGAGAGATC 800
GCGCGCGAGA TCGCGTGGAA GGGCCTCCCC GCCGCAGAGC GCGGTGAGAT 850
CCGTCTTGCC CTGCCAGCCG GCGTCGATGA GCAGGCGCCA GGTCGCGGGG 900
GTCGCCTGCA TCATCGTCGC TCTGCACGAT TCGATGCGCT CGCGAAGACG 950
CTCGCCGTCG AGCACGTCGC CGCGGGAGGC GATGACCGTC CTCCCGCCGA 1000
CGACGAGAGG CAAGAACAGC TCGAGACCCG CGATGTCGAA CGACGGCGTG 1050
GTGACCGCGA GGAGCACGTC GCCGGCTCGC AAGCCTGGCT CCTTCTGCAT 1100
GGCGCGCAGG AAATTCACGA GCTGGCGGTG CTCGATCTCG ACCCCCTTCG 1150

GCTTGCCCGT CGTGCCCGAC GTGTAGAT 1178

(2) INFORMATION FOR SEQ ID NO:92:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Val Glu Ile
5 10 15

Glu His Arg Gln Leu Val Asn Phe Leu Arg Ala Met Gln Lys Glu
20 25 30

Pro Gly Leu Arg Ala Gly Asp Val Leu Leu Ala Val Thr Thr Pro
35 40 45

Ser Phe Asp Ile Ala Gly Leu Glu Leu Phe Leu Pro Leu Val Val
50 55 60

Gly Gly Arg Thr Val Ile Ala Ser Arg Gly Asp Val Leu Asp Gly
65 70 75

Glu Arg Leu Arg Glu Arg Ile Glu Ser Cys Arg Ala Thr Met Met
80 85 90

Gln Ala Thr Pro Ala Thr Trp Arg Leu Leu Ile Asp Ala Gly Trp
95 100 105

Gln Gly Lys Thr Asp Leu Thr Ala Leu Cys Gly Gly Glu Ala Leu
110 115 120

Pro Arg Asp Leu Ala Arg Asp Leu Ser Ser Arg Val Arg Ala Leu
125 130 135

Trp Asn Met Tyr Gly Pro Thr Glu Thr Thr Ile Trp Ser Thr Val
140 145 150

His Gln Val Thr Asp Cys Asp Arg Asp Ile Pro Ile Gly His Pro
155 160 165

Ile Ala Asn Thr Asn Val Tyr Val Leu Asp Ala Ala Gly Arg Gln
170 175 180

Ala Pro Ile Gly Val Pro Gly Glu Leu Cys Ile Gly Gly Glu Gly
185 190 195

Val Ala Arg Gly Tyr Arg His Arg Pro Glu Leu Thr Ala Glu Lys
200 205 210

Phe Val Val Val Glu Val Ala Gly Arg Pro Ala Glu Arg Val Tyr
215 220 225

Arg Thr Gly Asp Ile Val Arg Leu Arg Ala Asp Arg Ala Leu Glu
230 235 240

Phe Val Gly Arg Arg Asp His Gln Val Lys Leu Arg Gly Tyr Arg
245 250 255

Ile Glu Leu Gly Glu Ile Glu Ser Val Leu Ser Glu His Pro Ser
260 265 270

Val Arg Arg Ser Val Ala Val Val Arg Glu Asp Thr Pro Gly Asp
275 288 285

Gln Arg Leu Val Ala Tyr Val Val Pro Ala Glu Asp Gly Val Leu
290 295 300

Pro Leu Asp Glu Leu Arg Thr Ala Leu Arg Ala Arg Leu Pro Glu
305 310 315

Tyr Met Val Pro Gln Val Ile Val Ser Leu Ala Ala Leu Pro Leu
320 325 330

Thr Pro Asn Asp Lys Ile Asp Arg Lys Ala Leu Pro Ile Pro Gln
335 340 345

Val Asn Arg Ala Ala Ala Ala Ala Ala Asp Thr Ala Ile Met Thr
350 355 360

Asp Val Gln Arg Gln Val Ala Ala Ile Trp Cys Ala Val Leu Gln
365 370 375

Leu Glu Arg Val Gly Leu His Glu Asn Phe Phe Ala Pro Gly Gly
380 385 390

Pro Ser

(2) INFORMATION FOR SEQ ID NO:93:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:1178
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

ATCTACACCT CCGGCACGAC GGGCAAGCCG AAGGGAGTAA AGATCACACA 50
TCGTGCCGTG GTGAATTTTC TGAACTCGAT GCGGCGTGAA CCAGGGCTGA 100
CCCCGGACGA TGTGGTGCTC TCGGTCACCA CGCTGTCGTT TGACATTGCC 150
GGACTCGAAC TCCACCTGCC CCTGACGACT GGAGCCACGG TCGTAGTGGC 200
GACCCAAGAC GCGGTGTCCG ACGCTGAACT GCTGGTCAGA GAGTTGGAGC 250

GGACCGGAAC AACTCTGTTG CAGGCGACGC CAGTCACATG GCGAATGCTT 300
CTGGAGTCGG GCTGGAAAGG AAATCCGCGA CTCAAGGCTC TGGTCGGAGG 350
TGAGGCAGTG CCGAGGGACC TGGTGAATCG GCTTGCTCCC CTTTGCGCGT 400
CACTTTGGAA CATGTACGGA CCAACGGAAA CCACGATCTG GTCAACGGTT 450
GGGCGTCTGG AGGCTGGAGA TGGTGTGTCT AGTATTGGCC GGCCCATCGA 500
CAATACGCGG ATTTACGTCG TGGATCCGTC GATACACCTT CAGCCCATCG 550
GAGTTCCCGG CGAATTGCTG ATTGGCGGAG AAGGATTGGC CGACGGATAT 600
CTGAAACGCG ATCAGTTGAC GGCAGAGAAG TTCATTCCTG ATCCATTTGG 650
TGGGAGGCCT GGGTCTCGGC TGTATCGAAC CGGAGATCTT GCGCGCTGGC 700
GCGCGGACGG CACCTTGGAG TGTCTCGGAC GAATGGACCA ACAGGTGAAG 750
ATTCGGGGTT CCCGGATCGA ATTGGGTGAG ATCGAAACCC TGTTGGCCTC 800
CCACCCGGAT GTGAAACAGA ACGTGGTGGT CGTACGCGAG GACAGCCCCG 850
GGGAAAAAAA ATTGGTGGGC TATTTCGTGC CGGCGAACGG ACGCAATCCC 900
GAAGTGATGG AATTTCGCAA ACATCTGCAG CGGACGCTTC CGGATTACAT 950
GGTCCCCTCA GTGTACGTGC CCTTGACCTC GGTTCCGCTT ACACCCAACG 1000
GAAAGATCGA CCGCAAGGCG CTGCCCGCAC CGGATATCAG CGCCGTCACG 1050
GTTTCCCGAG AGTCAATTGC GCCGCGCAAT CCCGCCGAAG AGCGGCTGGC 1100
AGCAATTTTC GCCAAGGTGC TTGGCACGCC GATCGCCTCG ATCCACGACA 1150
GCTTCTTCTC CCCGGGCGGC CCCTCCAT 1178

(2) INFORMATION FOR SEQ ID NO:94
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(iii) HYPOTHETICAL: no
(v) FRAGMENT TYPE: internal fragment
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Ile Tyr Thr Ser Gly Thr Thr Gly Lys Pro Lys Gly Val Lys Ile
5 10 15

Thr His Arg Ala Val Val Asn Phe Leu Asn Ser Met Arg Arg Glu
20 25 30

Pro Gly Leu Thr Pro Asp Asp Val Val Leu Ser Val Thr Thr Leu
35 40 45

Ser Phe Asp Ile Ala Gly Leu Glu Leu His Leu Pro Leu Thr Thr
50 55 60

Gly Ala Thr Val Val Val Ala Thr Gln Asp Ala Val Ser Asp Ala
65 70 75

Glu Leu Leu Val Arg Glu Leu Glu Arg Thr Gly Thr Thr Leu Leu
80 85 90

Gln Ala Thr Pro Val Thr Trp Arg Met Leu Leu Glu Ser Gly Trp
95 100 105

Lys Gly Asn Pro Arg Leu Lys Ala Leu Val Gly Gly Glu Ala Val
110 115 120

Pro Arg Asp Leu Val Asn Arg Leu Ala Pro Leu Cys Ala Ser Leu

125 130 135

Trp Asn Met Tyr Gly Pro Thr Glu Thr Thr Ile Trp Ser Thr Val
140 145 150

Gly Arg Leu Glu Ala Gly Asp Gly Val Ser Ser Ile Gly Arg Pro
155 160 165

Ile Asp Asn Thr Arg Ile Tyr Val Val Asp Pro Ser Ile His Leu
170 175 180

Gln Pro Ile Gly Val Pro Gly Glu Leu Leu Ile Gly Gly Glu Gly
185 190 195

Leu Ala Asp Gly Tyr Leu Lys Arg Asp Gln Leu Thr Ala Glu Lys
200 205 210

Phe Ile Pro Asp Pro Phe Gly Gly Arg Pro Gly Ser Arg Leu Tyr
215 220 225

Thr Gly Asp Leu Ala Arg Trp Arg Ala Asp Gly Thr Leu Glu
230 235 240

Cys Leu Gly Arg Met Asp Gln Gln Val Lys Ile Arg Gly Ser Arg
245 250 255

Glu Leu Gly Glu Ile Glu Thr Leu Leu Ala Ser His Pro Asp
260 265 270

Lys Gln Asn Val Val Val Val Arg Glu Asp Ser Pro Gly Glu
275 288 285

Lys Lys Leu Val Gly Tyr Phe Val Pro Ala Asn Gly Arg Asn Pro
290 295 300

Glu Val Met Glu Phe Arg Lys His Leu Gln Arg Thr Leu Pro Asp
305 310 315

Tyr Met Val Pro Ser Val Tyr Val Pro Leu Thr Ser Val Pro Leu
320 325 330

Thr Pro Asn Gly Lys Ile Asp Arg Lys Ala Leu Pro Ala Pro Asp
335 340 345

Ile Ser Ala Val Thr Val Ser Arg Glu Ser Ile Ala Pro Arg Asn
350 355 360

Pro Ala Glu Glu Arg Leu Ala Ala Ile Phe Ala Lys Val Leu Gly
365 370 375

Thr Pro Ile Ala Ser Ile His Asp Ser Phe Phe Ser Pro Gly Gly
380 385 390

Pro

CLAIMS

1. A method for recovery of antibiotic biosynthetic DNA from humic materials or lichen comprising the steps of:

(a) combining a humic or lichen-derived sample with a set of amplification primers under conditions suitable for polymerase chain reaction amplification, wherein the primer set is a degenerate primer set selected to hybridize with conserved regions of antibiotic biosynthetic gene;

(b) cycling the combined sample through a plurality of amplification cycles to amplify DNA complementary to the primer set; and

(c) isolating the amplified DNA.

2. The method according to claim 1, wherein the primer set hybridizes with a polyketide synthase gene.

3. The method according to claim 2, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 1 and 2.

4. The method according to claim 2, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 3 and 4.

5. The method according to claim 2, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 5 and 6.

6. The method according to claim 2, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 11 and 12.

7. The method according to claim 1, wherein the primer set hybridizes with a isopenicillin N synthase gene.

8. The method according to claim 7, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 7 and 8.

9. The method according to claim 1, wherein the primer set hybridizes with a peptide synthetase gene.

10. The method according to claim 9, wherein the primer set comprises primers having the sequence set forth in SEQ ID Nos. 9 and 10.

11. The method according to any of claims 1 to 10, wherein the sample comprises DNA extracted from a soil sample.

12. The method according to claim 1, wherein the sample is a lichen-derived sample.

13. The method according to any of claims 1 to 12, further comprising the steps of cloning the isolated DNA into a host organism, and isolating the cloned DNA.

14. The method according to claim 13, wherein the host organism is *E. coli*.

15. An oligonucleotide primer having the sequence as defined in any of Seq. ID. Nos. 1 through 8.

16. A composition comprising two oligonucleotide primers having the sequence as defined in Seq. ID Nos. 1 and 2; 3 and 4; 5 and 6; or 7 and 8.

17. A polynucleotide comprising a region having the sequence given by any of sequence ID Nos. 13, 15, 17, 19, 21, 23, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91 or 93.

18. A biosynthetic polypeptide encoded by a polynucleotide comprising a region having the sequence given by any of sequence ID Nos. 13, 15, 17, 19, 21, 23, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79 81, 83, 85, 87, 89, 91 or 93.

19. The biosynthetic polypeptide of claim 18, wherein the polypeptide has the amino acid sequence given by any of Sequence ID Nos. 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 3,4 3,6 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 or 94.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68 | A3

(11) International Publication Number: WO 98/53097

(43) International Publication Date: 26 November 1998 (26.11.98)

(21) International Application Number: PCT/CA98/00488

(22) International Filing Date: 21 May 1998 (21.05.98)

(30) Priority Data:
08/861,774 22 May 1997 (22.05.97) US

(71) Applicant: TERRAGEN DIVERSITY INC. [CA/CA]; University of British Columbia, Suite 300, 2386 East Mall, Vancouver, British Columbia V6T 1Z3 (CA).

(72) Inventors: WATERS, Barbara; 5706 Timbervalley Road, Delta, British Columbia V4L 2E6 (CA). MIAO, Vivian, P., W.; 13750 31 Avenue, Surrey, British Columbia V4P 2B7 (CA). YAP, Wai, Ho; 5 Elite Terrace, Singapore 458748 (SG). SEOW, Kah, Tong; 8 Jln Aneka, Serene Park, Johor Baru, Johor 80300 (MY).

(74) Agent: DEETH WILLIAMS WALL; National Bank Builing, Suite 400, 150 York Street, Toronto, Ontario M5H 3S5 (CA).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
11 March 1999 (11.03.99)

(54) Title: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR BIOACTIVE MOLECULES

(57) Abstract

Degenerate primers which hybridize with various classes of antibiotic biosynthesis gene were used to amplify fragments of DNA from soil and lichen extracts. Cloning and sequencing of the amplified products showed that these products included a variety of novel and previously uncharacterized antibiotic biosynthesis gene sequences, the products of which have the potential to be active as antibiotics, immunosuppressors, antitumor agents, etc. Thus, antibiotic biosynthesis genes can be recovered from soil or lichens by combining a sample with a pair of amplification primers under conditions suitable for polymerase chain reaction amplification, wherein the primer set is a degenerate primer set selected to hybridize with conserved regions of known antibiotic biosynthetic pathway genes, for example Type I and Type II polyketide synthase genes, isopenicillin N synthase genes, and peptide synthetase genes, cycling the combined sample through a plurality of amplification cycles to amplify DNA complementary to the primer set; and isolating the amplified DNA.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Internati' 'Application No
PCT/CA 98/00488

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97 12991 A (TERRAGEN DIVERSITY INC) 10 April 1997 see the whole document ---	1-14
Y	MALPARTIDA F. ET AL.,: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE, vol. 325, - 26 February 1987 pages 818-821, XP002075972 see the whole document ---	1-14
A	WO 87 03907 A (LUBRIZOL GENETICS INC) 2 July 1987 see the whole document --- -/--	1-14

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 August 1998	26. 01. 1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, F

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KATZ L ET AL: "POLYKETIDE SYNTHESIS: PROSPECTS FOR HYBRID ANTIBIOTICS" ANNUAL REVIEW OF MICROBIOLOGY, vol. 47, 1993, pages 875-912, XP000654850 see the whole document ---	1-14
A	CORTES J. ET AL.,: "An unusually large multifuntional polypeptide in the erythromycin producing polyketide synthase of Saccharopolyspora erythrarea" NATURE, vol. 348, - 8 November 1990 pages 176-178, XP002075973 see the whole document -----	1-14

1

INTERNATIONAL SEARCH REPORT

Interr. nal application No.
PCT/CA 98/ 00488

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see FURTHER INFORMATION SHEET

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-16 (complete)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-16 (complete)

   Invention 1:
   Method for recovering different polynucleotide species by using degenerated primers, primers and compositions therefore (Seq. Ids.: 1-12)

2. Claims 17-19 (complete)

   Invention 2:
   Biosynthetic polypeptides (amino acid sequences, nucleic acid sequences (and regions thereof) Seq. Ids.: 13 and 14.

   Inventions 3-42:
   ...ibidem for each sequence pair 15/16, 17-18 ...93/94 separately

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/CA 98/00488

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9712991 A	10-04-1997	AU 6922196 A	28-04-1997
		CA 2232709 A	10-04-1997
		EP 0851938 A	08-07-1998
WO 8703907 A	02-07-1987	AU 598516 B	28-06-1990
		AU 6835487 A	15-07-1987
		EP 0262154 A	06-04-1988
		EP 0463707 A	02-01-1992